INVERSE POINT SOURCE LOCATION WITH THE HELMHOLTZ EQUATION ON A BOUNDED DOMAIN

KONSTANTIN PIEPER, BAO QUOC TANG, PHILIP TRAUTMANN, AND DANIEL WALTER

ABSTRACT. The problem of recovering acoustic sources, more specifically monopoles, from point-wise measurements of the corresponding acoustic pressure at a limited number of frequencies is addressed. To this purpose, a family of sparse optimization problems in measure space in combination with the Helmholtz equation on a bounded domain is considered. A weighted norm with unbounded weight near the observation points is incorporated into the formulation. Optimality conditions and conditions for recovery in the small noise case are discussed, which motivates concrete choices of the weight. The numerical realization is based on an accelerated conditional gradient method in measure space and a finite element discretization.

Keywords: Inverse source location, Sparsity, Helmholtz equation, PDE-constrained optimization

1. Introduction

In this paper, we consider the problem of recovering a sound source u, consisting of an unknown number time-harmonic monopoles, from pointwise measurements of the acoustic pressure. It is well known that under the assumption of a time-harmonic signal consisting of N frequencies, the acoustic wave equation can be reduced to a family of Helmholtz equations. Concretely, let $\Omega \subset \mathbb{R}^d$, $d \in \{2,3\}$ be a bounded, convex, and polygonal (two dimensional) or polyhedral (three dimensional) domain. The boundary $\partial\Omega$ is partitioned into perfectly reflecting walls contained in $\Gamma_N \subset \partial\Omega$, and $\Gamma_Z = \partial\Omega \setminus \Gamma_N$ modeling absorbing walls or artificial boundaries arising from a truncation of an unbounded domain. We model the acoustic pressure $p_n \in L^2(\Omega)$ at the n-th frequency as the solution of

$$\begin{cases} -\Delta p_n - k_n^2 p_n = u_n|_\Omega & \text{in } \Omega, \\ \partial_\nu p_n - \mathrm{i}\kappa_n p_n = u_n|_{\Gamma_Z} & \text{on } \Gamma_Z, \\ \partial_\nu p_n = u_n|_{\Gamma_N} & \text{on } \Gamma_N, \end{cases} \tag{1.1}$$

where $n = 1, 2, \ldots, N$. Here, $k_n > 0$ is a sequence of wavenumbers, which are defined as usual by $k_n = \omega_n/c$, where c is the *speed of sound* and $\{\omega_n\}_n$ a set of *circular frequencies.* The numbers $\kappa_n \in \mathbb{C}$ with $\operatorname{Re}\kappa_n \neq 0$ are related to the properties of walls that are modeled on the boundary Γ_Z; cf. [2]. In the simplest case, we set $\kappa_n = k_n$, and obtain the well-known zeroth-order absorbing boundary

---

The authors gratefully acknowledge support through the International Research Training Group IGDK 1754, funded by the German Science Foundation (DFG) and the Austrian Science Fund (FWF). K. Pieper acknowledges funding by the US Department of Energy Office of Science grant DE-SC0016591 and by the US Air Force Office of Scientific Research grant FA9550-15-1-0001. D. Walter acknowledges support from the TopMath Graduate Center of TUM Graduate School and from the TopMath Program at the Elite Network of Bavaria.

conditions [19, 26]. We model the source u_n by a superposition of N_d acoustic monopoles,

$$u_n = \sum_{j=1}^{N_d} \boldsymbol{u}_{j,n} \delta_{\hat{x}_j}, \tag{1.2}$$

where $\boldsymbol{u}_{j,n} \in \mathbb{C}$ and $\hat{x}_j \in \Omega_c$, where $\Omega_c \subset \overline{\Omega}$ is a set containing all possible source locations. We suppose that for a finite number of observation points $\Xi = \{x_m \mid m = 1, \ldots, M\}$ pressure values $p_d^m \in \mathbb{C}^N$ of (1.1) are given (in the form of noisy recordings at M microphones, i.e. $p_d^m = p(x_m) + z^m$, $z^m \in \mathbb{C}^N$). Based on these observations the number of point sources N_d, the positions $\hat{x}_j \in \Omega_c$ and coefficients $\boldsymbol{u}_j \in \mathbb{C}^N$ are to be reconstructed. Inverse problems of this kind are of great importance in engineering applications such as beamforming [36, 39, 40, 41]. For instance, one is interested in locating a source of noise pollution using processed data captured by a microphone array.

Due to the fact that we have only partial observations of the acoustic pressure, the problem is under-determined, and therefore ill-posed. Thus we solve it based on a regularized least-squares formulation. We follow the approach of [6] and consider the following convex problem:

$$\begin{aligned} \min_{u \in \mathcal{M}_w(\Omega_c, \mathbb{C}^N)} \quad & \frac{1}{2} \sum_{m=1}^{M} |p(x_m) - p_d^m|_{\mathbb{C}^N}^2 + \alpha \|u\|_{\mathcal{M}_w(\Omega_c, \mathbb{C}^N)}, \\ \text{subject to} \quad & (1.1), \end{aligned} \tag{1.3}$$

where $p = (p_1, \ldots, p_N)$ and $u = (u_1, \ldots, u_N)$. In this problem, the solution of (1.3) is searched in the space of $\mathbb{C}^N$-valued Radon measures which satisfy

$$\|u\|_{\mathcal{M}_w(\Omega_c, \mathbb{C}^N)} = \int_{\Omega_c} |wu'|_{\mathbb{C}^N} \, \mathrm{d}|u| < \infty$$

for a vector-valued weighting function $w \colon \Omega_c \to \mathbb{C}^N$. Here, the point-wise product $w(x)u'(x) = (w_1(x)u_1'(x), \ldots, w_N(x)u_N'(x))$ should be understood in the sense of the Hadamard-product. The regularization functional promotes the sparsity of the support of the solution in Ω_c independent of the frequency components (also referred to as group or directional sparsity [24]); see [6, 32]. More concretely, it promotes solutions of the structure (1.2).

Note that, a more direct reconstruction approach would be the solution of the problem

$$\begin{aligned} \min_{x_j \in \Omega_c, \boldsymbol{u}_j \in \mathbb{C}^N} \quad & \frac{1}{2} \sum_{m=1}^{M} |p(x_m) - p_d^m|_{\mathbb{C}^N}^2 + \alpha \sum_{j=1}^{N_d} |w(x_j) \boldsymbol{u}_j|_{\mathbb{C}^N}, \\ \text{subject to} \quad & (1.1) \quad \text{with } u = (u_n)_n \text{ as in (1.2)}, \end{aligned} \tag{1.4}$$

where the number of sources N_d is fixed, but can be regarded as an additional discrete optimization variable. Since the locations x_j are now considered optimization variables, this is a non-convex finite-dimensional optimization problem with constraints $x_j \in \Omega_c$, which complicates the numerical solution. At first glance, the problem formulation (1.3) seems to be more general than (1.4) since we discard the structural assumption on the source u by considering general Borel measures. However, the existence of minimizers to (1.3) of the form (1.2) can be guaranteed for $N_d \leq 2NM$. Hence, if the number of sources N_d is left free, both problems are essentially equivalent, i.e. we can obtain a solution to the nonconvex problem (1.4) by solving the convex version (1.3).

The objective of this work is to provide a systematic theoretical development of the above recovery approach, including analysis of the problem, conditions for recovery, and algorithmic solution and numerical discretization strategies. In the case $w \equiv 1$ the analysis of the problem (1.3) relies on the assumption that the observation points and the control set Ω_c are separated from each other. However, by using weighting functions in the regularization functional with specific properties this restriction can be overcome. Moreover, an optimal choice of the weight function is shown to lead to improved theoretical and practical properties of the approach.

1.1. **Related works.** The analysis of the recovery approach is based on the analysis of the noise-free case, which leads to the corresponding minimum norm problem

$$\begin{aligned} \min_{u\in\mathcal{M}_w(\Omega_c,\mathbb{C}^N)} \quad & \|u\|_{\mathcal{M}_w(\Omega_c,\mathbb{C}^N)} \\ \text{subject to} \quad & p(x_m) = p_d^m, \quad m = 1,\dots,M, \end{aligned} \tag{1.5}$$

where p is the solution of (1.1); see, e.g., [6, 8, 17]. For $w \equiv 1$ it is shown in [6] that the solutions of (1.3) converge for $\alpha \to 0$ and $|z|^2_{\mathbb{C}^{MN}}/\alpha \to 0$ to a solution of (1.5) in the weak-star sense; see also [7, 25]. This can be carried over to the weighted case easily. We also note that the inverse problem under consideration can be interpreted as a deconvolution problem for measures involving the Green's function corresponding to (1.1) as convolution kernel. Problems of this form have been studied recently in [1, 8, 9, 17]. In [9] the recoverability of an exact source from convolutions with the Féjer kernel is proven under the assumption that the exact point sources are sufficiently well separated from each other. Concerning the use of a non-constant weight $w \neq 1$ we refer to [34]. By an appropriate choice of the weighting function the authors prove an exact recoverability result for a general deconvolution problem on a one-dimensional domain without requiring a minimum separation distance between the exact source points. However, these results are not directly applicable in our setting due to the more complicated structure of the convolution kernel under consideration.

Robustness with respect to noise has been investigated in [1, 8, 17]. In [17] it is shown that a strengthened source condition for small enough noise level δ and regularization parameter α the solution of (1.3) is unique and consists of the same number of point sources as the exact solution. Convergence rates for coefficients and positions of the reconstructed source to the exact coefficients and positions are derived.

Moreover, we mention that, after discretization on a finite grid, the inverse problem under consideration corresponds to an inverse problem involving an over-complete dictionary; see, e.g., [38]. The dictionary is given by point-evaluations of the Green's functions of (1.1). In the noise-free case such problems are often solved by a problem formulation corresponding to (1.5) (Basis Pursuit), and in the noisy case a problem corresponding to (1.3) is solved (LASSO). In most of the literature concerning over-complete dictionaries it is assumed that the entries of the dictionary have unit norm, in order to prevent bias in the dictionary. In our problem this is not the case. However, a particular form of the weight function

$w(x)$ leads to reweighted versions of the problems (1.3) and (1.5) in the variable $v = wu$, which have a dictionary with entries of unit norm.

Finally, concerning the discretization of the PDE-constrained optimization problem, a problem similar to (1.4) has been proposed [2] for a fixed number N_d and FE-discretizations have been analyzed in (cf. also [15]). Concerning the regularity and numerical analysis for sparse control problems with measures, in combination with different PDEs, we also refer to [10, 11, 27, 28].

1.2. **Contribution.** Concerning the analysis of (1.3), we first focus on the case $w \equiv 1$, which is complicated by the presence of point-wise sources (which lead to unbounded solutions) with point-wise observations of the solution. Nevertheless, based on regularity results for (1.1), we show that (1.3) and (1.5) are well-posed if the sources are restricted to some compact set Ω_c which does not contain the set of observation points Ξ. Note that this implies $\operatorname{dist}(\Omega_c, \Xi) > 0$. While this may not seem like a severe restriction, it introduces additional questions: On the one hand, a large distance restricts the possible location from where sources can be recovered. On the other hand, for a too small distance the problem favors sources close to the observation points, which introduces undesirable reconstruction artifacts. In fact, it can be proven that the problem with $w \equiv 1$ has no solutions if $\Xi \cap \Omega_c \neq \emptyset$; see Proposition 4.1. By introduction of a weight function w that is unbounded in the observation points, well-posedness of (1.3) can be shown for arbitrary Ω_c; see Section 4. Concerning the structure of the solutions, we show both problems always admit solutions of the form (1.2) with $N_d \leq 2NM$.

Clearly, not all sources of the form (1.2) can be recovered by (1.3). However, we show that all minimum norm solutions of (1.5) fulfill a *source condition*, which allows us to deduce convergence rates for the convergence of the solutions of (1.3) to solutions of (1.5) for vanishing noise and appropriately chosen α; see Section 5. Additionally, we give numerical examples of recoverable and non-recoverable sources. Even in the simple case of one unknown source, recoverability can fail unless an appropriate weight is employed. Moreover, numerical experiments suggest that the use of specific weights increases the number of recoverable sources. This is confirmed by statistical test involving randomly chosen positions and coefficients of the exact sources. In the case of a single point source we are able to prove that the exact source is the unique solution of (1.5) when using a specific weighting function and under additional assumption on the forward operator; see Proposition 5.9.

Concerning the numerical solution of (1.3), we adopt the algorithmic strategy proposed in [6] (see also [4]), which operates on the linear span of Dirac delta functions and combines point-insertion and removal steps. Moreover, a function space convergence theory is available, which bounds the number of necessary steps to obtain a prescribed accuracy in the functional value. We augment the procedure by an additional step which guarantees that the size of the support of the iterations of the algorithm can not grow beyond $2NM$. In [6] Dirac deltas are removed using one step of a proximal gradient method applied to (1.3) for the magnitudes with fixed positions. To further promote the sparsity of the iterates, this finite dimensional non-smooth optimization problem is resolved in every iteration (cf. also [4]) by means of a globalized semi-smooth Newton method. Additionally, we employ a discretization of (1.3) with finite elements for p and Dirac delta functions in the grid nodes. Although this transforms (1.3) into a finite dimensional optimization problem (amenable to a wide range of optimization algorithms), the function space analysis of the presented algorithm ensures that the number of iterations stays (uniformly) bounded for arbitrarily fine meshes.

This paper is organized in the following way. In Section 2 we establish regularity properties of the Helmholtz equation needed for the analysis of the optimization problem. Section 3 is devoted to the analysis of the problem with $w \equiv 1$. Section 4 is concerned with the weighted problem for a general weight. In Section 5, the regularization properties of the reconstruction procedure are investigated. Section 6 describes the optimization algorithm we use for the solution of the measure-valued optimal control problem. Finally, in Section 7 we conduct several numerical experiments.

1.3. **Notation and conventions.** Throughout the paper we adopt the following conventions: The complex numbers $\mathbb{C}$ are regarded as a $\mathbb{R}$-linear vector space endowed with the inner product $(z,v)_{\mathbb{C}} = \operatorname{Re}(z\bar{v}) = \operatorname{Re}(z)\operatorname{Re}(v) + \operatorname{Im}(z)\operatorname{Im}(v)$. Correspondingly, we denote the inner product on the Hilbert space $L^2(\Omega) = L^2(\Omega, \mathbb{C})$ by

$$(v,\varphi)_\Omega = \int_\Omega \operatorname{Re}(v\bar{\varphi})\,\mathrm{d}x.$$

This convention extends to all other inner products or duality pairings defined on derived spaces. We identify the space of $\mathbb{C}^N$-valued vector measures as

$$\mathcal{M}(\Omega_c, \mathbb{C})^N \cong \mathcal{M}(\Omega_c, \mathbb{C}^N) \cong \mathcal{C}(\Omega_c, \mathbb{C}^N)^*,$$

where the second isomorphism is isometric if $\mathcal{C}(\Omega_c, \mathbb{C}^N)$, the space of continuous functions with values in $\mathbb{C}^N$, is endowed with the norm $\|\varphi\|_{\mathcal{C}(\Omega_c,\mathbb{C}^N)} = \sup_{x\in\Omega_c}|\varphi(x)|_{\mathbb{C}^N}$. The duality pairing is defined by

$$\langle u, \varphi\rangle = \operatorname{Re}\left(\int_{\Omega_c} \bar{\varphi}\,\mathrm{d}u\right) = \int_{\Omega_c} (u', \varphi)_{\mathbb{C}^N}\,\mathrm{d}|u| = \sum_{n=1}^N \operatorname{Re}\left(\int_{\Omega_c} \bar{\varphi}_n\,\mathrm{d}u_n\right),$$

with the total variation measure $|u| \in \mathcal{M}^+(\Omega_c)$ (in the space of positive Borel measures), the Radon-Nikodym derivative $u' = \mathrm{d}u/\,\mathrm{d}|u| \in L^1(\Omega_c, \mathbb{C}^N, \mathrm{d}|u|)$, and $u_n \in \mathcal{M}(\Omega_c)$ the signed real valued measures arising as the component measures of u. By C we denote a generic constant, which has different values at different appearances.

2. Analysis of the Helmholtz equation

Let $\Omega \subset \mathbb{R}^d$, $d \in \{2,3\}$ be a bounded, convex, and polytopal domain. Following [2], we assume that the boundary is of the form $\Gamma = \partial\Omega = \Gamma_N \cup \Gamma_Z$ where $\Gamma_N = \cup_m \overline{\Gamma_m}$ can be written as the union of some subset of plane faces of Γ and that $\Gamma_Z = \partial\Omega \setminus \Gamma_N$. We note that these assumption on the boundary could be relaxed considerably, at the expense of making the following arguments more technical; see Remark 1 below. For simplicity, we follow the setting of [2]. Moreover, we assume that Γ_Z has positive measure, which is needed to ensure unique solvability for all wave numbers. We denote the characteristic function of Γ_Z by $\chi_{\Gamma_Z} : \Gamma \to \{0,1\}$.

Denote by $\Omega_c \subset \overline{\Omega}$ the control set, which is required to be closed (and therefore compact). The state equation problem reads as: find $p = (p_1, \ldots, p_N)$ for $n \in$

$\{1,2,\dots,N\}$ where $p_n\colon \Omega \to \mathbb{C}$ solves

$$\begin{cases} -\Delta p_n - k_n^2 p_n = u_n|_\Omega & \text{in } \Omega, \\ \partial_\nu p_n - \mathrm{i}\kappa_n \chi_{\Gamma_Z} p_n = u_n|_\Gamma & \text{on } \Gamma, \end{cases} \tag{2.1}$$

$k_n > 0$ are real numbers and $u = (u_1, u_2, \dots, u_N) \in \mathcal{M}(\Omega_c, \mathbb{C}^N)$ is a vector measure. Note that, in the interest of generality, we allow the measure to be supported on the boundary. These contributions of the measure appear in the boundary conditions, but are included in the weak formulation given below in a natural way.

In this section, we assume without restriction that $N = 1$ and suppress the dependency on n of k, κ, u and p. The general case of the results follows directly from the (complex) scalar case.

Definition 1 (Very weak solutions for (2.1))**.** Let $u \in \mathcal{M}(\bar{\Omega})$ be a complex valued measure. A complex valued function $p \in L^2(\Omega)$ is said to be a solution *by transposition* to (2.1) if it satisfies

$$(p,q)_\Omega = \langle u, r\rangle \quad \text{for all } q \in L^2(\Omega), \tag{2.2}$$

where $r \in H^2(\Omega)$ is the solution to the dual problem

$$\begin{cases} -\Delta r - k^2 r = q & \text{in } \Omega, \\ \partial_\nu r + \mathrm{i}\bar{\kappa}\chi_{\Gamma_Z} r = 0, & \text{on } \Gamma. \end{cases} \tag{2.3}$$

Note, that the duality pairing $\langle u, r\rangle$ is well defined due to the continuous embedding $H^2(\Omega) \hookrightarrow \mathcal{C}(\overline{\Omega})$ for spatial dimension $d \le 3$. It can be shown that the solution by transposition also satisfies the following *very weak* formulation:

$$-(p, \Delta\varphi + k^2\varphi)_\Omega = \langle u, \varphi\rangle$$
$$\text{for all } \varphi \in H^2(\Omega) \text{ with } \partial_\nu\varphi + \mathrm{i}\bar{\kappa}\chi_{\Gamma_Z}\varphi = 0 \text{ on } \Gamma. \tag{2.4}$$

Theorem 2.1. *For any $u \in \mathcal{M}(\Omega_c)$, there exists a unique very weak solution $p \in L^2(\Omega)$ to* (2.1) *and there holds*

$$\|p\|_{L^2(\Omega)} \le C\|u\|_{\mathcal{M}(\Omega_c)}.$$

Proof. This result is proven by the method of transposition as in Definition 1 (cf. [29]) using the $H^2(\Omega)$-regularity of the unique solution of the dual equation (2.3); see [2, Theorem 3.3]. For the underlying regularity theory for the Neumann problem on convex polytopal domains we refer also to [13, 23]. □

Lemma 2.2. *The very weak solution $p \in L^2(\Omega)$ from Theorem 2.1 has the improved regularity $p \in W^{1,s}(\Omega)$ for any $s < d/(d-1)$ and there holds*

$$\|p\|_{W^{1,s}(\Omega)} \le C\|u\|_{\mathcal{M}(\Omega_c)}.$$

Proof. This result can be proved by using a Hölder continuity result for the dual equation (2.3) with weak formulation

$$(\nabla\varphi, \nabla r)_\Omega - (k^2\varphi, r)_\Omega - (\mathrm{i}\bar{\kappa}\chi_{\Gamma_Z}\varphi, r)_\Gamma = \langle q, \varphi\rangle_{W^{-1,s'}(\overline{\Omega}), W^{1,s}(\Omega)}$$

with data $q \in W^{-1,s'}(\overline{\Omega}) = (W^{1,s}(\Omega))^*$, i.e., with $1/s' + 1/s = 1$ and the corresponding a priori estimate

$$\|r\|_{\mathcal{C}(\overline{\Omega})} \le C\|q\|_{W^{-1,s'}(\overline{\Omega})}.$$

Such a result can be found, e.g., in [22] (cf. also [16]). To apply the result, which is derived for real systems of equations, we split the solution into real and imaginary part, apply [22, Theorem 7.1 (i)], and use the embedding properties of Sobolev-Campanato spaces; see, e.g., [22, Theorem 2.1 (i)]. □

Based on the previous existence and regularity results, certain observations of the state solution p (e.g., in $L^2(\Omega)$ or $L^s(\Gamma)$ for $s \leq d/(d-1)$) are possible. To obtain the continuity of point evaluations, we use the smoothness of the solution away from the support of the source u. First we analyze the fundamental solutions.

Lemma 2.3. *Let $y \in \overline{\Omega}$. Then the very weak solution G^y to the equation*

$$\begin{cases} -\Delta G^y - k^2 G^y = \delta_y|_\Omega, & \text{in } \Omega, \\ \partial_\nu G^y - \mathrm{i}\kappa\chi_{\Gamma_Z} G^y = \delta_y|_\Gamma, & \text{on } \Gamma, \end{cases} \tag{2.5}$$

satisfies for $\varepsilon > 0$ the estimate

$$\|G^y\|_{H^2(\Omega\setminus\overline{B_\varepsilon(y)})} \leq C(\varepsilon), \tag{2.6}$$

where $B_\varepsilon(y)$ is the ε-ball around y, and C depends continuously on ε.

Proof. We follow standard arguments based on a smoothed indicator function. For completeness, we give a short sketch of the proof. Multiply G^y with a weight function $\zeta_\varepsilon \in \mathcal{C}_c^\infty(\overline{\Omega} \setminus \overline{B_{\varepsilon/2}(y)})$, such that $\zeta_\varepsilon(x) = 1$ for $x \in \overline{\Omega} \setminus \overline{B_\varepsilon}(y)$. Now, by the chain rule and (2.5), the product $G^y_\zeta = \zeta_\varepsilon G^y$ fulfills

$$\begin{cases} -\Delta G^y_\zeta - k^2 G^y_\zeta = -\Delta\zeta_\varepsilon G^y - 2\nabla\zeta_\varepsilon \nabla G^y, & \text{in } \Omega, \\ \partial_\nu G^y_\zeta - \mathrm{i}\kappa\chi_{\Gamma_Z} G^y_\zeta = G^y \partial_\nu \zeta_\varepsilon - \mathrm{i}\kappa\chi_{\Gamma_Z} G^y \zeta_\varepsilon, & \text{on } \Gamma. \end{cases} \tag{2.7}$$

Now, we use the facts that $G^y \in L^2(\Omega)$ with Theorem 2.1 and $\nabla G^y \in L^s(\Omega)$ for $s < d/(d-1)$ arbitrary with Lemma 2.2. With the trace theorem it additionally follows $G^y|_\Gamma \in L^s(\Gamma)$. By the Sobolev embedding in dimensions $d \leq 3$, we obtain $\nabla G^y \in H^{-1}(\Omega)$ (choose $s > 2d/(d+2)$) and $G^y|_\Gamma \in H^{-1/2}(\Gamma)$ (choose $s > 2-2/d$). Together with $\|\nabla^2\zeta_\varepsilon\|_{L^\infty(\bar{\Omega})} \leq C\varepsilon^{-2}$ it follows now from a classical result for (2.7) that $G^y_\zeta \in H^1(\Omega)$ with $\|G^y_\zeta\|_{H^1(\Omega)} \leq C/\varepsilon^2$. By the trace theorem, it follows that $G^y_\zeta|_\Gamma \in H^{1/2}(\Gamma)$. Now, we introduce $G^y_{\zeta^2} = \zeta_\varepsilon G^y_\zeta$ and repeat the argument to derive regularity of $G^y_{\zeta^2}$ from the previous results for G^y_ζ. By a H^2 regularity result (see, e.g., [2, Theorem 3.3]), we obtain $G^y_{\zeta^2} \in H^2(\Omega)$, with norm bounded by C/ε^{-4}. Since by construction $G^y_{\zeta^2}(x) = G^y(x)$ for all $x \in \Omega$ with $|x-y| \geq \varepsilon$, we obtain (2.6). □

Lemma 2.4. *Let $\mathcal{N}_\varepsilon(\Omega_c) = \{x \in \Omega \mid \operatorname{dist}(x, \Omega_c) < \varepsilon\}$. The solution p to* (2.1) *belongs to $\mathcal{C}(\overline{\Omega} \setminus \mathcal{N}_\varepsilon(\Omega_c))$ for all $\varepsilon > 0$ together with*

$$\|p\|_{\mathcal{C}(\overline{\Omega}\setminus\mathcal{N}_\varepsilon(\Omega_c))} \leq C(\varepsilon)\|u\|_{\mathcal{M}(\Omega_c)}.$$

Proof. We approximate u by a sequence of finite sum of Dirac delta measures, i.e., there exists a sequence $u_K \rightharpoonup^* u$ in $\mathcal{M}(\Omega_c)$ with $\|u_K\|_{\mathcal{M}(\Omega_c)} \leq \|u\|_{\mathcal{M}(\Omega_c)}$ and

$$u_K = \sum_{k=1}^K \boldsymbol{u}_k \delta_{y_k}$$

with $\boldsymbol{u}_k \in \mathbb{C}$ and $y_k \in \Omega_c$. By linearity, we have for the unique solution p_K of (2.1) corresponding to u_K that $p_K = \sum_{k=1}^K \boldsymbol{u}_k G^{y_k}$ where G^{y_k} is the solution of (2.5) with δ_{y_k} in place of δ_y. For every $\varepsilon > 0$ there exists a $C = C(\varepsilon)$ with

$$\|p_K\|_{L^2(\Omega)} + \|p_K\|_{H^2(\Omega\setminus\overline{\mathcal{N}_\varepsilon}(\Omega_c))} \le C\sum_{k=1}^K |\boldsymbol{u}_k| = C\|u_K\|_{\mathcal{M}(\Omega_c)} \le C\|u\|_{\mathcal{M}(\Omega_c)}$$

using Theorem 2.1 and Lemma 2.3. Hence, there exists a function $p \in L^2(\Omega) \cap H^2(\Omega \setminus \overline{\mathcal{N}_\varepsilon}(\Omega_c))$ such that

$$p_K \rightharpoonup p \quad \text{in} \quad L^2(\Omega) \cap H^2(\Omega \setminus \overline{\mathcal{N}_\varepsilon}(\Omega_c))$$

up to a subsequence. Using this weak convergence and $u_K \rightharpoonup^* u$ in $\mathcal{M}(\Omega_c)$ we can pass to the limit $K \to \infty$ to obtain that p is the very weak solution to the problem (2.1) and the estimate

$$\|p\|_{L^2(\Omega)} + \|p\|_{H^2(\Omega\setminus\overline{\mathcal{N}_\varepsilon}(\Omega_c))} \le C(\varepsilon)\|u\|_{\mathcal{M}(\Omega)}.$$

holds for some $C(\varepsilon) > 0$. Thus, the proof is complete when we use the embedding $H^2(\Omega \setminus \overline{\mathcal{N}_\varepsilon}(\Omega_c)) \hookrightarrow \mathcal{C}(\overline{\Omega} \setminus \mathcal{N}_\varepsilon(\Omega_c))$ for dimensions $d \le 3$. □

Clearly, the same regularity results also hold for the dual equation,

$$\begin{cases} -\Delta \bar{G}^y - k^2 \bar{G}^y = \delta_y, & \text{in } \Omega, \\ \partial_\nu \bar{G}^y + \mathrm{i}\bar{\kappa}\chi_{\Gamma_Z}\bar{G}^y = 0, & \text{on } \Gamma. \end{cases} \tag{2.8}$$

Note that the only difference between (2.5) and (2.8) occurs in the boundary conditions on Γ_Z. It is therefore easy to see that the solutions to (2.5) are (2.8) are the same up to complex conjugation, which justifies the notation $\bar{G}^y$. In the case $y \in \Omega$ (and not on Γ), we can give a more precise description of the nature of the singularity. We will need this for the adjoint equation in section 4.

Proposition 2.5. *Let $y \in \Omega$. Then the very weak solution $\bar{G}^y$ to the dual equation* (2.8) *can be written as $\bar{G}^y(x) = \bar{\Phi}^y(x) + \bar{\xi}^y(x)$ for $x \in \overline{\Omega}$, where*

$$\Phi^y(x) = \phi_k(|x-y|) = \begin{cases} (i/4)H_0^{(1)}(k|x-y|) & \text{for } d = 2, \\ \exp(\mathrm{i}k|x-y|)/(4\pi|x-y|) & \text{for } d = 3, \end{cases} \tag{2.9}$$

is a fundamental solution of the free space Helmholtz equation

$$-\Delta\Phi^y - k^2\Phi^y = \delta_y, \quad x \in \mathbb{R}^n, \tag{2.10}$$

(fulfilling the Sommerfeld radiation condition), and $\xi^y \in H^2(\Omega)$ is the solution to (2.11)*. The special function $H_0^{(1)}$ is the Hankel function of the first kind; see, e.g., [12, Section 3.4].*

Proof. We follow [2]. First, we consider a fundamental solution Φ^y to the Helmholtz equation in the whole domain (2.10). In fact Φ^y can be written explicitly as in (2.9); see, e.g., [12]. We will use the facts that $\Phi^y \in C^\infty(\mathbb{R}^n \setminus \{y\})$ and $\|\Phi^y\|_{\mathcal{C}^1(K)} \le C(\operatorname{dist}(y,K))|K|$ for any $K \subset\subset \Omega$. Then $\bar{G}^y$ is a solution of (2.8) if and only if $\bar{G}^y = \bar{\Phi}^y + \bar{\xi}^y$, with ξ^y satisfying

$$\begin{cases} -\Delta\xi^y - k^2\xi^y = 0, & \text{in } \Omega, \\ \partial_\nu\xi^y - \mathrm{i}\kappa\chi_{\Gamma_Z}\xi^y = -\partial_\nu\Phi^y + \mathrm{i}\kappa\chi_{\Gamma_Z}\Phi^y, & \text{on } \Gamma. \end{cases} \tag{2.11}$$

We have the following estimate for ξ^y (see, e.g., [2, Theorem 3.3]):

$$\|\xi^y\|_{H^2(\Omega)} \le C\left(\|\partial_\nu \Phi^y - \mathrm{i}\kappa\Phi^y\|_{H^{1/2}(\Gamma_Z)} + \|\partial_\nu \Phi^y\|_{H^{1/2}(\Gamma_N)}\right).$$

Thus, it follows directly $\|\xi^y\|_{H^2(\Omega)} \le C(\operatorname{dist}(y,\Gamma))$. □

Remark 1. The H^2 regularity of G^y in Lemma 2.3 (and of ξ^y in Proposition 2.5) uses the structural assumption on the polygonal domain, namely that the boundary conditions can only change on different plane faces of the boundary (based on the results in [2]). It is possible to relax this assumption, and consider more general domains Ω in two or three dimensions. We will comment on two possible options, which we however do not pursue here for the sake of brevity.

Hölder-regularity: By using the regularity results from, e.g., [16, 22] (as in Lemma 2.2), which are valid for much more general configurations of the boundary, we can get continuous solutions without H^2 regularity. The solution by transposition can be based on these regularity results directly; cf. [35, 37]. Additionally, Lemma 2.3 can be modified to show local Hölder-continuity, which again leads to the result of Lemma 2.4. A similar comment applies to Proposition 2.5.

Interior regularity: If we introduce a $\Omega' \subset\subset \Omega$, we can show alternative to Lemma 2.3 the result $G^y \in H^2(\Omega' \setminus \overline{B_\varepsilon}(y))$ without using any assumptions on the boundary beyond Lipschitz-continuity. The proof can be done as in Lemma 2.3, by suitably modifying the smoothed indicator function. For interior regularity results of elliptic equations cf. also [31, Theorem 47.1] [20, Theorems 9.11 and 9.13]. However, interior results do not allow to include point sources or pointwise observations on the boundary of the domain.

3. Analysis of the optimization problem

We suppose that for some points $\{x_m\}_{m=1,2,\dots,M} \subset \overline{\Omega} \setminus \Omega_c$ the acoustic pressure values $p_d^m \in \mathbb{C}^N$ are given. We consider the following optimization problem:

$$\min_{u\in\mathcal{M}(\Omega_c,\mathbb{C}^N)} J(p,u) = \frac{1}{2}\sum_{m=1}^{M} |p(x_m) - p_d^m|^2_{\mathbb{C}^N} + \alpha\|u\|_{\mathcal{M}(\Omega_c,\mathbb{C}^N)}, \tag{3.1}$$

$$\text{subject to} \quad \begin{cases} -\Delta p_n - k_n^2 p_n = u_n|_\Omega, & \text{in } \Omega, \\ \partial_\nu p_n - \mathrm{i}\kappa_n \chi_{\Gamma_Z} p_n = u_n|_\Gamma, & \text{on } \Gamma, \end{cases} \quad n = 1,2,\dots,N. \tag{3.2}$$

Since $x_m \notin \Omega_c$, there exists $\varepsilon_0 > 0$ such that $x_m \notin \mathcal{N}_{\varepsilon_0}(\Omega_c)$ for all $m = 1,2,\dots,M$. Due to Lemma 2.4 we can evaluate p_n at x_k and thus define the control-to-observation operator

$$S : \mathcal{M}(\Omega_c,\mathbb{C}^N) \to (\mathbb{C}^N)^M \quad \text{as} \quad Su = (p(x_1), p(x_2), \dots, p(x_M)).$$

We introduce the reduced optimal control problem

$$\min_{u\in\mathcal{M}(\Omega_c,\mathbb{C}^N)} j(u) = \frac{1}{2}\sum_{m=1}^{M} |(Su)_m - p_d^m|^2_{\mathbb{C}^N} + \alpha\|u\|_{\mathcal{M}(\Omega_c,\mathbb{C}^N)}, \tag{P_α}$$

which is clearly equivalent to (3.1)–(3.2).

We will see that S can alternatively be defined as the dual of a linear bounded operator S^*, to be introduced below.

Lemma 3.1. *$u_n \rightharpoonup^* u$ in $\mathcal{M}(\Omega_c,\mathbb{C}^N)$ implies $Su_n \to Su$ in $\mathbb{C}^{NM}$.*

By established arguments, we obtain the following basic existence result.

Proposition 3.2. *The problem* (P_α) *has an optimal solution* $\widehat{u}$.

To derive optimality conditions, we consider the adjoint equation,

$$\begin{cases} -\Delta\xi_n - k_n^2\xi_n = \sum_{\{m \mid x_m\in\Omega\}} q_{n,m}\delta_{x_m}, & \text{in } \Omega, \\ \partial_\nu\xi_n + \mathrm{i}\bar{\kappa}_n\chi_{\Gamma_Z}\xi_n = \sum_{\{m \mid x_m\in\Gamma\}} q_{n,m}\delta_{x_m}, & \text{on } \Gamma, \end{cases} \quad n = 1,2,\ldots,N, \tag{3.3}$$

for given $q \in \mathbb{C}^{NM}$. We denote the by S^* the operator that maps a given q to the restriction $\xi|_{\Omega_c}$, where $\xi = (\xi_1, \ldots, \xi_N)$ is the corresponding solution to (3.3).

Proposition 3.3. *The linear operator* $S^*\colon \mathbb{C}^{NM} \to \mathcal{C}$ *is bounded.*

Proof. First we note that the equation (3.3) has a measure right-hand side. However, since $x_m \in \Omega \setminus \mathcal{N}_{\varepsilon_0}(\Omega_c)$ for all $m = 1, 2, \ldots, M$, we have $\xi_n \in \mathcal{C}(\mathcal{N}_{\varepsilon_0}(\Omega_c)) \subset \mathcal{C}(\Omega_c)$ thanks to Lemma 2.3. Thus the operator S^* is well defined. The linearity of S^* is trivial. The boundedness of S^* follows with linearity from Lemma 2.3. ☐

Proposition 3.4. *The operator* S *is the dual of the operator* S^*, *that is*

$$(Su, q) = \langle u, S^*q\rangle = \sum_{m=1}^{M}\sum_{n=1}^{N}\langle u_n, \bar{G}_n^{x_m} q_{n,m}\rangle \tag{3.4}$$

for all $q \in \mathbb{C}^{NM}$ *and all* $u \in \mathcal{M}(\Omega_c, \mathbb{C}^N)$, *where* $\bar{G}_n^{x_m}$ *is defined in* (2.8) *with* $k = k_n$.

Proof. Similar to Lemma 2.4, we approximate u by a sequence u_K of the form $u_K = \sum_{k=1,\ldots,K} \boldsymbol{u}_k\delta_{y_k}$. From [2, Theorem 7.2], with a slight modification, we have for all K that

$$(Su_K, q) = \langle u_K, S^*q\rangle.$$

Passing to the limit as $K \to \infty$ and using Lemma 3.1 and $u_K \rightharpoonup^* u$ we get the desired result. The last equality in (3.4) follows by linearity of S^*. ☐

As in [6], the following optimality conditions system can be derived.

Proposition 3.5. *A measure* $\widehat{u} \in \mathcal{M}(\Omega_c, \mathbb{C}^N)$ *is a solution to* (P_α) *if and only if* $\widehat{\xi} = -S^*(S\widehat{u} - p_d)$ *satisfies* $\|\widehat{\xi}\|_{\mathcal{C}(\Omega_c,\mathbb{C}^N)} \le \alpha$ *and the polar decomposition* $\mathrm{d}\widehat{u} = \widehat{u}'\,\mathrm{d}|\widehat{u}|$, *with* $\widehat{u}' \in L^1(\Omega_c, |\widehat{u}|, \mathbb{C}^N)$, *satisfies*

$$\alpha\widehat{u}' = \widehat{\xi} \qquad |\widehat{u}|\text{-almost everywhere.}$$

Thereby, $\operatorname{supp}|\widehat{u}| \subset \{x \in \Omega_c \mid |\widehat{\xi}(x)|_{\mathbb{C}^N} = \alpha\}$ *for each solution* $\widehat{u}$.

Proof. The proof follows the one of [6, Proposition 3.6] with minor modification concerning the complex valued measure and the compact control domain. ☐

Since the operator S maps into a finite dimensional space, the solution set of (P_α) always contains linear combinations of Dirac delta function. This can be seen by interpreting the corresponding dual problem as a *semi-infinite* optimization problem; see, e.g., [3, Section 5.4]. For the convenience of the reader, we provide an independent exposition in Appendix B.

Corollary 3.6. *There exists an optimal solution* $\widehat{u}$ *to* (P_α) *which consists of* $N_d \le 2NM$ *point sources,*

$$\widehat{u} = \sum_{j=1}^{N_d} \widehat{\boldsymbol{u}}_j\delta_{\widehat{x}_j} \quad \text{where } \widehat{\boldsymbol{u}}_j \in \mathbb{C}^N,\ \widehat{x}_j \in \Omega_c.$$

Proof. This follows by combining Proposition B.3 with Theorem B.4. Note that it holds $\dim \operatorname{Ran} S \leq \dim \mathbb{C}^{NM} = 2NM$, since $\mathbb{C}$ is regarded as a real vector space. □

Corollary 3.7. *Any solution* $\widehat{u} = \sum_{j=1}^{N_d} \widehat{\boldsymbol{u}}_j \delta_{\widehat{x}_j}$ *from Corollary 3.6 is uniquely characterized by the optimality conditions*

$$\|\widehat{\xi}\|_{\mathcal{C}(\Omega_c,\mathbb{C}^N)} \leq \alpha, \qquad \alpha\, \widehat{\boldsymbol{u}}_j = |\widehat{\boldsymbol{u}}_j|_{\mathbb{C}^N} \widehat{\xi}(\widehat{x}_j), \quad j \in \{1, 2, \ldots, N_d\},$$

where $\widehat{\xi} = -S^*(S\widehat{u} - p_d)$ *is the associated adjoint state.*

4. Weighted norm approach

In practical computations, the recovery based on (P_α) succeeds only in some cases. In particular, there exist single point-sources which can not be recovered even in the noise-free case. These cases occur when the boundary of the set Ω_c is close to the observation points (in which case several spurious sources tend to be placed in these spots), or if the exact source is located in a spot with "bad" acoustical properties; see section 7. Consider for a moment the case $N = 1$, and assume that the exact source is given by $u^\star = \boldsymbol{u}^\star \delta_{x^\star}$. The magnitude of the observed signal is given by

$$|Su^\star|_{\mathbb{C}^M} = |\boldsymbol{u}^\star| \sqrt{\sum_{m=1}^{M} |G^{x_m}(x^\star)|^2} = |\boldsymbol{u}^\star| \hat{w}(x^\star)$$

Thus, the magnitude of the observation for a unit source originating from $x \in \Omega$ is described by the function $\hat{w}\colon \Omega \to \mathbb{R}_+ \cup \{+\infty\}$. Empirically, the cases of non-identifiability coincide with the cases where $\hat{w}(x^\star)$ is small, compared to a global value such as, e.g., $\max_{x\in\Omega_c} \hat{w}(x)$ or the mean of $\hat{w}$. However, if the magnitude of each source is computed in the weighted norm,

$$\|u^\star\|_{\mathcal{M}_{\hat{w}}(\Omega_c,C^N)} = \int_{\Omega_c} \hat{w}\, \mathrm{d}|u^\star| = |\boldsymbol{u}^\star| \hat{w}(x^\star),$$

a source of unit size leads to an observation of unit size.

Motivated by this, we introduce for each frequency n a weight w^n and consider a weighted problem:

$$\min_{u\in\mathcal{M}_w(\Omega_c,\mathbb{C}^N)} J_w(p,u) = \frac{1}{2} \sum_{m=1}^{M} |p(x_m) - p_d^m|^2_{\mathbb{C}^N} + \alpha \|u\|_{\mathcal{M}_w(\Omega_c,\mathbb{C}^N)}, \quad \text{subject to} \quad (3.2). \tag{4.1}$$

In the interest of generality, we consider a formulation with a general class of weights. We will define the weighted norm $\|\cdot\|_{\mathcal{M}_w(\Omega,\mathbb{C}^N)}$ for admissible choices of the weight w below.

In a weighted problem formulation, the technical condition on the observation points $x_m \notin \Omega_c$ can be avoided. Therefore, in the following, we only assume that $\Omega_c \subset \Omega$ is closed in Ω. Let $\Xi = \{x_m \mid m = 1, 2, \ldots, M\} \subset \Omega$ be the observation points (pairwise distinct). For simplicity, we do not consider boundary observation in this section. Note that the original problem (3.1)–(3.2) is not necessarily well-posed in such cases.

Proposition 4.1. *Suppose that Ω_c does not contain isolated points and that* $\operatorname{dist}(\Xi, \Omega_c) = 0$. *Then, without restriction, $x_m \in \Omega_c$ for $1 \le m \le M_1 \le M$ and $x_m \notin \Omega_c$ for $m > M_1$. If $(p_d^m)_{m=1,\dots,M_1} \in \mathbb{C}^{NM_1}$ is sufficiently large (or $M_1 = M$),* (3.1)–(3.2) *does not admit a solution.*

Proof. For simplicity of notation, we assume without restriction that $N = 1$. Denote the optimization problem (3.1)–(3.2) by (P_{orig}). Consider first a modified optimization problem, where we minimize

$$J_{\text{aux}}(p,u) = \frac{1}{2}\sum_{m=M_1+1}^{M} |p(x_m) - p_d^m|_{\mathbb{C}}^2 + \alpha\|u\|_{\mathcal{M}(\Omega_c,\mathbb{C})},$$

subject to (3.2). We denote the corresponding optimization problem by (P_{aux}). By similar arguments as in section 3, there exists an optimal solution $u_0 \in \mathcal{M}(\Omega_c, \mathbb{C})$ to the modified problem (P_{aux}). By optimality, we obtain that

$$\|u_0\|_{\mathcal{M}(\Omega_c,\mathbb{C})} \le \frac{\|(p_d^m)_{m=M_1+1,\dots,M}\|_{\mathbb{C}^{M-M_1}}^2}{2\alpha}.$$

By continuity, it holds $\|Su_0\|_{\mathbb{C}^M} \le C\|(p_d^m)_{m=M_1+1,\dots,M}\|_{\mathbb{C}^{M-M_1}}^2$ for a generic $C > 0$ and any solution of (P_{aux}). Clearly, $\min(P_{\text{aux}}) \le \inf(P_{\text{orig}})$. In fact, equality holds: We show that for

$$u^n = u_0 + \sum_{m=1}^{M_1} \boldsymbol{u}_m^n \delta_{x_m^n}, \tag{4.2}$$

with appropriate $\boldsymbol{u}_m^n \to 0 \in \mathbb{C}$, $x_m^n \to x_m$ it holds $J(Su^n, u^n) \to \min(P_{\text{aux}})$. To this purpose, we first fix $x_m^n \in \Omega_c$ with $|x_m - x_m^n| = r_m^n$, for $r_m^n > 0$ with $r_m^n \to 0$ as $n \to \infty$. Then, we consider the matrix $M^n \in \mathbb{C}^{M_1 \times M_1}$, which results from the restriction of S to the span of $\delta_{x_m^n}$ in the domain space and to the first M_1 observations in the image space, that is

$$M_{m,k}^n = G^{x_k^n}(x_m) \quad \text{for } m,k = 1,\dots,M_1.$$

Moreover, recalling the definition of ϕ_k, see (2.9), we introduce the diagonal matrix

$$D^n = \operatorname{diag}\left(\frac{1}{|\phi_{k_1}(r_1^n)|}, \dots, \frac{1}{|\phi_{k_1}(r_{M_1}^n)|}\right).$$

By Proposition 2.5 and the properties of the Green's functions, we derive that

$$D^n M^n \to \mathrm{Id}_{\mathbb{C}^{M_1}} \quad \text{for } n \to \infty.$$

Thus we have $|\det(D^n)||\det(M^n)| = |\det(D^n M^n)| > 1/2$ for n large enough. Consequently, for n large enough the matrix M^n is invertible. We can therefore choose $\boldsymbol{u}^n = (\boldsymbol{u}_1^n, \dots, \boldsymbol{u}_{M_1}^n)$ to be the solution of the system of equations $(M^n \boldsymbol{u}^n)_m = p_d^m - (Su_0)_m$ for $m = 1,\dots,M_1$. Therefore we have $(S(u^n))_m = p_d^m$ for $m = 1,\dots,M_1$, thanks to (4.2), and since $|\phi_{k_1}(r_m^n)| \to \infty$ for $n \to \infty$, it follows additionally that $\boldsymbol{u}_m^n \to 0$ for $m = 1,\dots,M_1$. This shows that $u^n \to u_0$ strongly in $\mathcal{M}(\Omega_c, \mathbb{C}^N)$ and $\inf(P_{\text{orig}}) \le J(Su^n, u^n) \to \min(P_{\text{aux}}) \le \inf(P_{\text{orig}})$ for $n \to \infty$. Assume now that (P_{orig}) admits a solution $\widehat{u}$. With $J(S\widehat{u}, \widehat{u}) = \inf(P_{\text{orig}}) = \min(P_{\text{aux}})$ we immediately deduce that $(S\widehat{u})_m = p_d^m$ for $m = 1,\dots,M_1$, and $\widehat{u}$ also solves (P_α). However, choosing $\|(p_d^m)_{m=1,\dots,M_1}\|_{\mathbb{C}^{M_1}}$ large enough contradicts the bound $\|S\widehat{u}\|_{\mathbb{C}^M} \le C\|(p_d^m)_{m=M_1+1,\dots,M}\|_{\mathbb{C}^{M-M_1}}^2$ which follows from the optimality of $\widehat{u}$ for (P_{aux}). □

Now, we introduce the class of admissible weight functions.

Definition 2 (Admissible weights)**.** We call a family of weight functions $w^n\colon \Omega_c \to \mathbb{R}\cup\{+\infty\}$, $n\in\{1,2,\dots,N\}$ admissible, if they fulfill the following properties:

i) $\inf_{x\in\overline{\Omega}} w^n(x) > 0$,

ii) w^n is upper semi-continuous and w^n restricted to $\Omega_c\setminus\Xi$ is continuous.

iii) The function $G_n^{x_m}/w^n$ can be continuously extended from $\Omega_c\setminus\Xi$ to Ω_c.

For admissible weights, we denote $[G_n^{x_m}/w^n](x_m) = \lim_{x\to x_m} G_n^{x_m}(x)/w^n(x)$. The case $[G_n^{x_m}/w^n](x_m) = 0$ for all m is of special interest.

Due to the fact that $|G_n^{x_m}(x)|\to\infty$ for $x\to x_m$, the upper semi-continuity of w^n and Property iii) imply that $w^n(x_m)=+\infty$. Now, we construct functions w^n such that the above conditions hold. With regard to the representation formula from Lemma 2.3, we can take for instance the functions

$$w^n_{\text{free}} = \sum_{m=1}^{M} |\varPhi_n^{x_m}| \tag{4.3}$$

In the following, we will again suppress the dependency on n, for convenience of notation.

Proposition 4.2. *The weights given in* (4.3) *are admissible.*

Proof. Property i) holds by the properties of the Green's functions. In both the two- and three-dimensional case, the functions $|\varPhi^0(x)|$ are radially symmetric and monotonously decreasing towards zero for $|x|\to\infty$. Therefore, $|\varPhi^{x_m}(x)| = |\varPhi^0(x-x_m)|$ is uniformly bounded from below on Ω for all m. By a similar argument, property ii) follows. It remains to verify iii). With Lemma 2.3, we notice that

$$\frac{G^{x_m}(x)}{w_{\text{free}}(x)} = \frac{\xi^{x_m}(x)+\varPhi^{x_m}(x)}{w_{\text{free}}(x)} = \frac{\xi^{x_m}(x)}{w_{\text{free}}(x)} + \frac{\varPhi^{x_m}(x)}{w_{\text{free}}(x)}$$

with $\xi^{x_m}\in H^2(\Omega)$. Since $\inf_{x\in\overline{\Omega}} w_{\text{free}}(x) > 0$ and for all points $\hat{x}$ where w_{free} is discontinuous it holds $\lim_{x\to\hat{x}} w_{\text{free}}(x) = +\infty$, the first term is continuous and we have

$$\lim_{x\to x_m}\frac{\xi^{x_m}(x)}{w_{\text{free}}(x)} = 0.$$

Furthermore, w_{free} has the form $w_{\text{free}}(x) = f_m(x) + |\varPhi^{x_m}(x)|$ for an $f_m\colon\Omega\to\mathbb{R}_+\cup\{+\infty\}$, which is finite and continuous in a neighborhood of x_m. Thus we have

$$\lim_{x\to x_m}\frac{\varPhi^{x_m}(x)}{w_{\text{free}}(x)} = \lim_{x\to x_m}\frac{\varPhi^{x_m}(x)}{|\varPhi^{x_m}(x)|} = 1.$$

In fact, for this, we use the concrete formulas for $\varPhi^{x_m}$; see Lemma 2.3. In the case $d=3$, it holds that $\varPhi^{x_m}(x) = \exp(\mathrm{i}k|x-x_m|)/4\pi|x-x_m|$, and the equality follows directly. In the case $d=2$, we use that for $t=k|x-x_m|$ we have

$$\varPhi^{x_m}(x) = \frac{i}{4}H_0^{(1)}(t) = -\frac{1}{4}Y_0(t) + \frac{i}{4}J_0(t),$$

where $J_0\colon\mathbb{R}_+\to\mathbb{R}$ and $Y_0\colon\mathbb{R}_+\to\mathbb{R}$ are the Bessel functions of the first and second kind. It is known that J_0 is continuous at $t=0$ and Y_0 is diverging towards $+\infty$ at $t=0$; see, e.g., [12, Section 3.4]. □

Remark 2. We verify that w^n_{free} is independent of the wave number k_n in three dimensions, since $|\Phi^{x_m}_n(x)| = 1/(4\pi|x - x_m|)$. In two dimensions, the singularity of $|\Phi^{x_m}_n|$ is of same type as the singularity of the Green's function of the Laplacian, $g(x) = -1/(2\pi)\ln|x-x_m|$, and k_n enters only in an additive constant; see, e.g., [12, Section 3.4]. Therefore, we could alternatively take the same weight for all n.

Other families of weight functions can be based on the Green's function on the domain. For instance, they are given by

$$w^n_{\Omega,1} = \sum_{m=1}^{M} |G^{x_m}_n|, \qquad w^n_{\Omega,2} = \sqrt{\sum_{m=1}^{M} |G^{x_m}_n|^2}. \tag{4.4}$$

Note that these weights depend on the shape of Ω and the wave number k_n. As for (4.3), we obtain the admissibility of (4.4).

Proposition 4.3. *Suppose that for any n there exists no $x \in \bar{\Omega}$, such that $G^{x_m}_n(x) = 0$ for all m. Then, the weights given in (4.4) are admissible.*

Proof. With Lemma 2.3, the verification of ii) and iii) follows by straightforward computations, since the local behavior of w_{free} and $w^n_{\Omega,1}$, $w^n_{\Omega,2}$ at the observation points are the same. For the uniform boundedness from below it suffices to observe that $w^n(x) > 0$ for all $x \in \Omega_c \setminus \Xi$, the w^n are continuous on the same set, and $w^n(x) \to \infty$ for $x \to x_m$. □

Remark 3. Certainly, there are many more possibilities to define admissible weights. For instance, we can use a different discrete norm for the absolute values of the Green's functions associated with the x_m or employ a weighed sum. Moreover, the weight for each m could be used as a separate regularization parameter, to obtain a more flexible regularization strategy.

For any vectors $v, w \in \mathbb{C}^N$, we define by $vw \in \mathbb{C}^N$ the coordinate-wise, or Hadamard product. Define now the weighted norm

$$\|u\|_{M_w(\Omega_c,\mathbb{C}^N)} = \int_{\Omega_c} |wu'| \,\mathrm{d}|u| = \int_{\Omega_c} \sqrt{\sum_{n=1}^{N} (w^n(x)|u'_n(x)|)^2} \,\mathrm{d}|u|(x)$$

Since $u' \in L^\infty(\Omega_c, |u|, \mathbb{C}^N)$ and w is upper semi-continuous, the function under the integral is positive and Borel-measurable, and the integral is well-defined for any $u \in \mathcal{M}(\Omega_c, \mathbb{C}^N)$ (but not necessarily finite). Note that if $w^n = w$ for all n, we obtain the more intuitive form

$$\|u\|_{\mathcal{M}_w(\Omega_c,\mathbb{C}^N)} = \int_{\Omega_c} w \,\mathrm{d}|u|.$$

We define the corresponding subspace of $\mathcal{M}(\Omega_c, \mathbb{C}^N)$ as

$$\mathcal{M}_w(\Omega_c, \mathbb{C}^N) = \left\{ u \in \mathcal{M}(\Omega_c, \mathbb{C}^N) \;\middle|\; \int_{\Omega_c} |wu'| \,\mathrm{d}|u| < \infty \right\}.$$

Next, we introduce the mapping $W \colon \mathcal{M}(\Omega_c, \mathbb{C}^N) \to \mathcal{M}_w(\Omega_c, \mathbb{C}^N)$ defined by

$$\mathrm{d}W(v) = \frac{v'}{w} \,\mathrm{d}|v|.$$

Again, the division v/w for $v, w \in \mathbb{C}^N$ is understood in a coordinate-wise fashion. We adopt the convention $z/(+\infty) = 0$ for any $z \in \mathbb{C}$.

Proposition 4.4. *Let w fulfill property i) and ii) and $w(\Xi \cap \Omega_c) \equiv +\infty$. The mapping W is well-defined and surjective. Moreover, the restriction*

$$W|_{\mathcal{M}(\Omega_c \setminus \Xi, \mathbb{C}^N)} \colon \mathcal{M}(\Omega_c \setminus \Xi, \mathbb{C}^N) \to \mathcal{M}_w(\Omega_c, \mathbb{C}^N)$$

is an isometric isomorphism.

Proof. The function $x \mapsto 1/w(x)$ is continuous on Ω_c according to the assumptions. Thus $W(v)$ is an element of $\mathcal{M}(\Omega_c, \mathbb{C}^N)$. Trivially, there holds $W(v) \in \mathcal{M}_w(\Omega_c, \mathbb{C}^N)$ for any $v \in \mathcal{M}(\Omega_c, \mathbb{C}^N)$. Additionally, for any $u \in \mathcal{M}_w(\Omega_c, \mathbb{C}^N)$, the product uw defined by

$$\mathrm{d}(uw) = wu' \,\mathrm{d}|u|$$

gives an element in $\mathcal{M}(\Omega_c, \mathbb{C}^N)$ since w is upper semi-continuous. Clearly, we have $W(uw) = u$ and thus W is surjective. However, W is not injective, and the kernel of W can be characterized as

$$\ker W = \mathcal{M}(\Xi, \mathbb{C}^N) = \left\{ \sum_{m=1}^{M} u_m \delta_{x_m} \;\middle|\; u_m \in \mathbb{C}^N \right\}$$

In fact, let v be an element of $\ker W$. Thus there holds $\|W(v)\|_{\mathcal{M}(\Omega_c, \mathbb{C}^N)} = \int_{\Omega_c} 1/w \,\mathrm{d}|v| = 0$, which is equivalent to

$$\operatorname{supp} v = \operatorname{supp}|v| \subseteq \{x \in \Omega_c \mid 1/w(x) = 0\} = \Xi \cap \Omega_c.$$

As a direct consequence of the isomorphism theorem, we obtain that

$$W \colon \mathcal{M}(\Omega_c, \mathbb{C}^N)/\ker W \to \mathcal{M}_w(\Omega_c, \mathbb{C}^N)$$

is an isomorphism. It can be directly verified that the quotient space is isomorphic to $\mathcal{M}(\Omega_c \setminus \Xi, \mathbb{C}^N)$; see, e.g., [33, Theorem 4.9 a)]. ☐

Based on these observations, we transform the weighted problem to one with weight one, which enables us to reuse the general results. We introduce a new optimization variable $v = uw \in \mathcal{M}(\Omega_c, \mathbb{C}^N)$ and employ a reduced formulation in terms of v. The corresponding observation operator and its adjoint are defined as

$$(S^w v, q) = \langle v, (S^w)^* q \rangle = \sum_{n=1}^{N} \sum_{m=1}^{M} \langle v, (\bar{G}_n^{x_m} / w_n)\, q_{n,m} \rangle, \tag{4.5}$$

for any $v \in \mathcal{M}(\Omega_c, \mathbb{C}^N), q \in \mathbb{C}^{NM}$. For any admissible weight, due to property iii), this yields a well defined operator.

Proposition 4.5. *For any admissible w, the operators $S^w \colon \mathcal{M}(\Omega_c, \mathbb{C}^N) \to \mathbb{C}^{NM}$ and $(S^w)^* \colon \mathbb{C}^{NM} \to \mathcal{C}(\Omega_c, \mathbb{C}^N)$ are well-defined and continuous with respect to the weak-$*$ topology and bounded, respectively.*

Now, we consider the reduced optimization problem

$$\min_{v \in \mathcal{M}(\Omega_c, \mathbb{C}^N)} j_w(v) = \frac{1}{2} \sum_{m=1}^{M} |(S^w v)_m - p_d^m|_{\mathbb{C}^N}^2 + \alpha \|v\|_{\mathcal{M}(\Omega_c, \mathbb{C}^N)}. \tag{$P_{\alpha,w}$}$$

Since the reweighed problem $(P_{\alpha,w})$ has exactly the same structural properties as the reduced problem (P_α), all results from sections 3 and 5 can be transferred without modification. In particular, for any admissible weight the problem $(P_{\alpha,w})$ admits optimal solutions $\widehat{v} \in \mathcal{M}(\Omega_c, \mathbb{C}^N)$ consisting of at most $2NM$ Dirac delta functions.

Given a solution $\widehat{v}$ of $(P_{\alpha,w})$ which does not contain any Dirac delta functions in the observation points (i.e., $\widehat{v} \in \mathcal{M}(\Omega_c \setminus \Xi, \mathbb{C}^N)$), we can apply W to obtain a solution of the original problem. First, we need some result to connect the algebraically defined operator S^w to the point evaluations of the solutions of (3.2).

Lemma 4.6. *For $\varepsilon > 0$ define $\Xi^\varepsilon = \cup_{m=1,\dots,M} B_\varepsilon(x_m)$. Let the observation operator $S_\varepsilon\colon \mathcal{M}(\Omega_c \setminus \Xi^\varepsilon, \mathbb{C}^N) \to \mathbb{C}^{NM}$ be defined as $S_\varepsilon(u) = (p(x_m))_{m=1,\dots,M}$, where p is the solution to* (3.2) *(defined with Lemma 2.4).*

If w is admissible, the operator $S^w\colon \mathcal{M}(\Omega_c \setminus \Xi, \mathbb{C}^N) \to \mathbb{C}^{NM}$ is the unique weak-$*$ continuous extension of the family of operators $S^w_\varepsilon = S_\varepsilon \circ W$.

Proof. By a simple computation, $S^w|_{\mathcal{M}(\Omega_c \setminus \Xi, \mathbb{C}^N)}$ extends all S^w_ε and by Proposition 4.5 it is continuous. Clearly, the spaces $\mathcal{M}(\Omega_c \setminus \Xi^\varepsilon, \mathbb{C}^N)$ are weak-$*$ dense in $\mathcal{M}(\Omega_c \setminus \Xi, \mathbb{C}^N)$, which proves the uniqueness of the extension. □

Lemma 4.7. *Let w be admissible.*

i) Suppose that $(P_{\alpha,w})$ possesses a solution $\widehat{v}$ with $\operatorname{supp} \widehat{v} \subset \Omega_c \setminus \Xi$*. Then, $\widehat{u} = W\widehat{v}$ is a solution of* (4.1)*.*

ii) Conversely, suppose that any solution of $(P_{\alpha,w})$ fulfills $|\widehat{v}|(\Xi) > 0$ and that Ω_c contains no isolated points. Then (4.1) *possesses no solution.*

Proof. Based on Proposition 4.5 and Lemma 4.6, the point evaluations of the solutions to (3.2) with sources in $\mathcal{M}_w(\Omega_c, \mathbb{C}^N)$ are well-defined. Moreover, using the isometric isomorphism property of W from Proposition 4.5, the infimum of (4.1) is equal to

$$\widehat{\jmath} = \inf_{v \in \mathcal{M}(\Omega_c \setminus \Xi, \mathbb{C}^N)} j_w(v). \tag{4.6}$$

Clearly, the minimum of $(P_{\alpha,w})$ fulfills $\min_{v \in \mathcal{M}(\Omega_c, \mathbb{C}^N)} j_w(v) \le \widehat{\jmath}$.

Now, if $(P_{\alpha,w})$ admits a solution $\widehat{v} \in \mathcal{M}(\Omega_c \setminus \Xi, \mathbb{C}^N)$, it follows that $j_w(\widehat{v}) = \widehat{\jmath}$ and the infimum of (4.1) is assumed by $\widehat{u} = W\widehat{v}$.

Conversely, if any solution to $(P_{\alpha,w})$ is not in $\mathcal{M}(\Omega_c \setminus \Xi, \mathbb{C}^N)$, the infimum in (4.6) is not assumed. To see this, we first show that it in fact holds that $\min_{v \in \mathcal{M}(\Omega_c, \mathbb{C}^N)} j_w(v) = \widehat{\jmath}$. Take any sparse solution $\widehat{v}$ of $(P_{\alpha,w})$. By the assumption, it contains Dirac delta functions supported on Ξ. Since the support points which coincide with observation points are not isolated in Ω_c, we can slightly perturb them, such that $x_m \neq \tilde{x}^\ell_m \to x_m$ for $\ell \to \infty$. Denote the perturbed measure by $\tilde{v}_\ell$. It holds $\|\tilde{v}_\ell\|_{\mathcal{M}(\Omega_c, \mathbb{C}^N)} = \|\widehat{v}\|_{\mathcal{M}(\Omega_c, \mathbb{C}^N)}$ and $\tilde{v}_\ell \in \mathcal{M}(\Omega_c \setminus \Xi, \mathbb{C}^N)$ for ℓ big enough and with the weak-$*$ continuity of S^w we obtain $\widehat{\jmath} \le \lim_{n \to \infty} j_w(\tilde{v}_\ell) = j_w(\widehat{v})$. Therefore, j_w can not assume its minimum on $\mathcal{M}(\Omega_c \setminus \Xi, \mathbb{C}^N)$, which directly implies that (4.1) has no minimum, using again Proposition 4.5 and Lemma 4.6. □

To obtain well-posedness of the weighted problem (4.1) without any assumptions on the structure of the solutions of the auxiliary problem $(P_{\alpha,w})$, we can impose the additional condition $[G^{x_m}/w](x_m) = 0$ for all m. For instance, for any admissible weight w (such as given in (4.3) or (4.4)) and some monotonously increasing function $\psi\colon \mathbb{R} \to \mathbb{R}_+$ with $\psi(0) = 0$, $\psi(t) > 0$ for $t > 0$, and $\psi(t)/t \to \infty$ for $t \to \infty$, the weight $\tilde{w} = \psi \circ w$ has this property.

Proposition 4.8. *Suppose that w is admissible with $[G^{x_m}/w](x_m) = 0$ for all m. Then, the operator S^w is weak-$*$ continuous on the space $\mathcal{M}(\Omega_c \setminus \Xi, \mathbb{C}^N)$.*

Proof. This follows directly from the observation that

$$(S^w)^* \colon \mathbb{C}^N \to \mathcal{C}_0(\Omega_c \setminus \Xi, \mathbb{C}^N) = \{v \in \mathcal{C}(\Omega_c, \mathbb{C}^N) \mid v(x_m) = 0, x_m \in \Xi \cap \Omega_c\}.$$

and the identification $\mathcal{C}_0(\Omega_c \setminus \Xi, \mathbb{C}^N)^* = \mathcal{M}(\Omega_c \setminus \Xi, \mathbb{C}^N)$. □

In this case, the solutions of $(P_{\alpha,w})$ are always supported on $\Omega_c \setminus \Xi$, which follows from the optimality conditions and the fact that $\widehat{\xi} = -(S^w)^*(S^w(\widehat{v}) - p_d)$ fulfills $\widehat{\xi}(\Xi \cap \Omega_c) = 0$. We summarize all results in the following theorem.

Theorem 4.9. *Let w be admissible and suppose that $(P_{\alpha,w})$ admits solutions in the space $\mathcal{M}(\Omega_c \setminus \Xi, \mathbb{C}^N)$ or that $[G^{x_m}/w](x_m) = 0$ for all m. Then, the problem (4.1) has a minimum $\widehat{u} \in \mathcal{M}_w(\Omega_c, \mathbb{C}^N)$ which consists of finitely many Dirac delta functions, $\widehat{u} = \sum_{j=1}^{N_d} \widehat{\boldsymbol{u}}_j \delta_{\widehat{x}_j}$. Together with the associated*

$$\widehat{\xi} = -S^*(S\widehat{u} - p_d),$$

it is uniquely characterized by the optimality conditions

$$\|\widehat{\xi}/w\|_{\mathcal{C}(\Omega_c,\mathbb{C}^N)} \leq \alpha, \qquad \alpha\, w(\widehat{x}_j)\widehat{\boldsymbol{u}}_j = |w(\widehat{x}_j)\widehat{\boldsymbol{u}}_j|_{\mathbb{C}^N}\, \widehat{\xi}(\widehat{x}_j)/w(\widehat{x}_j),$$

$j \in \{1, 2, \ldots, N_d\}$. Moreover, $\operatorname{supp}|\widehat{u}| \subset \{x \in \Omega_c \mid |\widehat{\xi}(x)/w(x)|_{\mathbb{C}^N} = \alpha\}$ *for each solution $\widehat{u}$.*

5. Regularization properties

In this section, we study (loosely speaking) if the minimization problem delivers an appropriate solution for the inverse problem: solve $Su = p$ for u. We mainly rely on general results for nonsmooth Tikhonov regularization [7, 25] and sparse spike deconvolution [6, 17]. To that purpose, we assume that we are given the exact source $u^\star$ of the form

$$u^\star = \sum_{j=1}^{N^\star} \boldsymbol{u}_j^\star \delta_{x_j^\star}, \quad \text{where } \boldsymbol{u}_j^\star \in \mathbb{C}^N \setminus \{0\},\ x_j^\star \in \Omega_c \setminus \Xi \tag{5.1}$$

and noisy observations $p_d = Su^\star + f = p^\star + f$ with small noise $\|f\|_{\mathbb{C}^{NM}} \leq \delta$. In the following we state conditions on $u^\star$ and a parameter choice rule for α in dependence of δ which are sufficient for the convergence of the solutions $\widehat{u}_\alpha$ of (3.1)–(3.2) (or the weighted problem (4.1)) towards the exact solution $u^\star$ for vanishing noise $\delta \to 0$ and for $\alpha(\delta) \to 0$. Moreover, convergence rates are given.

Without loss of generality, we only study the reduced weighted problem $(P_{\alpha,w})$ for a general admissible weight w. The case of $w \equiv 1$ with $\Omega_c \cap \Xi = \emptyset$ from section 3 is then included as a simple special case. In the case of solutions $\widehat{v}_\alpha$ of formulation $(P_{\alpha,w})$, we are interested in the convergence of $W\widehat{v}_\alpha$ towards $u^\star$. We define

$$v^\star = \sum_{j=1}^{N^\star} \boldsymbol{v}_j^\star \delta_{x_j^\star}, \quad \text{where } \boldsymbol{v}_j^\star = w(x_j^\star)\boldsymbol{u}_j^\star, \tag{5.2}$$

In the following, we study the convergence of solutions $\widehat{v}_{\alpha(\delta)}$ towards $v^\star$. Clearly, since $1/w$ is a continuous function on Ω_c, this implies convergence of $W\widehat{v}_\alpha$ towards $Wv^\star = u^\star$. We first analyse the following minimum norm problem, (cf., e.g., [6, 17, 25]):

$$\min_{v \in \mathcal{M}(\Omega_c,\mathbb{C}^N)} \|v\|_{\mathcal{M}(\Omega_c,\mathbb{C}^N)} \quad \text{subject to } S^w v = p^\star. \tag{$P_{0,w}$}$$

By assumption, the admissible set of $(P_{0,w})$ is not empty, since $p^\star = S^w v^\star$. Therefore, with Lemma 3.1, we can derive the following basic result; see Appendix B.

Proposition 5.1. *There exists a solution $v^\dagger \in \mathcal{M}(\Omega_c, \mathbb{C}^N)$ to $(P_{0,w})$, which consists of $N_d \leq 2NM$ point sources,*

$$v^\dagger = \sum_{j=1}^{N_d} \boldsymbol{v}_j^\dagger \delta_{x_j^\dagger} \quad \textit{where } \boldsymbol{v}_j^\dagger \in \mathbb{C}^N,\ x_j^\dagger \in \Omega_c.$$

We now turn to the limiting behavior of $(P_{\alpha,w})$ for small α and δ. From [25] (cf. [6, Section 4]), we have the following result.

Theorem 5.2. *For any monotonously increasing parameter choice rule $\alpha(\delta)$ for which $\delta^2/\alpha(\delta) \to 0$ and $\alpha(\delta) \to 0$ for $\delta \to 0$, any sequence $\widehat{v}_{\alpha(\delta)}$ of solutions to $(P_{\alpha,w})$ contain a subsequence which converges towards a solution $v^\dagger$ of $(P_{0,w})$ (weakly-$*$ in $\mathcal{M}(\Omega_c, \mathbb{C}^N)$). If additionally $v^\dagger$ is unique, the whole sequence converges towards $v^\dagger$.*

Under a *source condition* convergence rates can be derived in a generalized Bregman distance (see, e.g., [7]). It has the following form:

$$\text{There exists a } y^\dagger \in \mathbb{C}^{NM}, \text{ such that } (S^w)^* y^\dagger \in \partial \|v^\star\|_{\mathcal{M}(\Omega_c,\mathbb{C}^N)}. \tag{5.3}$$

A concrete form of this condition can be given by using the characterization of the subdifferential.

Proposition 5.3. *The source condition* (5.3) *can be equivalently expressed as: There exists a $y^\dagger \in \mathbb{C}^{NM}$, such that the associated adjoint state $\xi^\dagger = S^* y^\dagger$ fulfills*

$$\|\xi^\dagger/w\|_{\mathcal{C}(\Omega_c,\mathbb{C}^N)} \leq 1, \qquad \boldsymbol{v}_j^\star = |\boldsymbol{v}_j^\star|_{\mathbb{C}^N}\, \xi^\dagger(x_j^\star)/w(x_j^\star), \quad j \in \{1, 2, \ldots, N^\star\}.$$

The last condition can also be given by $w(x_j^\star)\boldsymbol{u}_j^\star = |w(x_j^\star)\boldsymbol{u}_j^\star|_{\mathbb{C}^N}\, \xi^\dagger(x_j^\star)/w(x_j^\star)$.

In our situation, the source condition is satisfied if $v^\star$ is a minimum norm solution, since (5.3) is a necessary and sufficient optimality condition of the minimum norm problem problem $(P_{0,w})$.

Proposition 5.4. *For any solution $v^\dagger$ of $(P_{0,w})$ there exists a corresponding $y^\dagger$ such that $(S^w)^* y^\dagger \in \partial\|v^\dagger\|_{\mathcal{M}(\Omega_c,\mathbb{C}^N)}$. Conversely, for any pair $y^\dagger$ and $v^\star$ fulfilling* (5.3)*, $v^\star$ is a solution of $(P_{0,w})$.*

Proof. This result follows by an application of Fenchel duality (see Propositions A.1 and A.2 in the Appendix). □

Corollary 5.5. *The element $v^\star$ satisfies the source condition* (5.3) *if and only if $v^\star$ is a solution of the minimum norm problem $(P_{0,w})$.*

Remark 4. The equivalence between the minimum norm problem and the source condition is due the semi-infinite character of the dual problem of (P_α); see Appendix A. In an general setting (with infinite dimensional observation) this equivalence is not always given; cf. [25].

The convergence rates for the regularized solutions will now be given in terms of a generalized, set-valued Bregman distance $D\colon \mathcal{M}(\Omega_c, \mathbb{C}^N) \times \mathcal{M}(\Omega_c, \mathbb{C}^N) \to \mathbb{R}$ defined by

$$D(v_1, v_2) = \{\|v_1\|_{\mathcal{M}(\Omega_c,\mathbb{C}^N)} - \|v_2\|_{\mathcal{M}(\Omega_c,\mathbb{C}^N)} - \langle \xi, v_1 - v_2\rangle \mid \xi \in \partial\|v_2\|_{\mathcal{M}(\Omega_c,\mathbb{C}^N)}\}$$

for any $v_1, v_2 \in \mathcal{M}(\Omega_c, \mathbb{C}^N)$. In [7, Theorem 2] (cf. also [6, Section 4]) the following convergence result is proven.

Theorem 5.6. *Let the source condition* (5.3) *be satisfied and let $\delta/c_1 \leq \alpha(\delta) \leq c_1\delta$ for some fixed $c_1 \geq 1$. Then for each minimizer $\widehat{u}_\alpha$ of* (3.1) *there exists a $d \in D(\widehat{u}_\alpha, u^\star)$ such that $d \leq C\delta$ holds (for some generic constant C).*

Based on Theorem 5.2 and Proposition 5.4 we see that the only missing part for the convergence of $\widehat{v}_{\alpha(\delta)}$ to $v^\star$ is the uniqueness of the solution of the minimum norm problem. Due to Proposition 5.1 unique solutions must necessarily consist of finitely many Dirac delta functions. Additionally, criteria for uniqueness based on the source condition can be derived. We give without proof the following popular one; cf. [14, Lemma 1.1] or [17, Proposition 5]:

Proposition 5.7. *Let $v^\star \in \mathcal{M}(\Omega_c, \mathbb{C}^N)$ with $v^\star = \sum_{j=1,\dots,N_d} \boldsymbol{v}_j^\star \delta_{x_j^\star}$, where $\boldsymbol{v}_j \in \mathbb{C}^N \setminus \{0\}$, $x_j^\star \in \Omega_c \setminus \Xi$ pairwise different. Suppose further that the source condition* (5.3) *holds with $\xi^\dagger/w = (S^w)^* y^\dagger$, the vectors $z_j = S^w(\xi_w^\dagger(x_j^\star)\delta_{x_j^\star}) \in \mathbb{C}^{MN}$ form a $\mathbb{R}$-linearly independent set, and for every $x \in \Omega_c \setminus \{x_j^\star \mid j = 1, 2, \dots, N^\star\}$ there holds $\|\xi^\dagger(x)/w(x)\|_{\mathbb{C}^{NM}} < 1$. Then $v^\star$ is the unique solution of $(P_{0,w})$.*

Finally, we sum up the findings of this section.

Corollary 5.8. *Let $u^\star = Wv^\star$, where $v^\star$ is a solution of $(P_{0,w})$ (or equivalently let $v^\star$ satisfy the source condition* (5.3)*) and let the conditions from Proposition 5.7 be satisfied. Furthermore, let $\delta/c_1 \leq \alpha(\delta) \leq c_1\delta$ for some $c_1 > 0$ as $\delta \to 0$. Then for any sequence of minimizers $\widehat{u}_{\alpha(\delta)}$ of $(P_{\alpha,w})$ it holds*

$$W\widehat{v}_{\alpha(\delta)} \rightharpoonup^* u^\star,$$

and there exists $d \in D(\widehat{v}_{\alpha(\delta)}, v^\star)$ such that $d \leq C\delta$ (for some generic C).

Due to the complex geometrical setup of (1.1) (in the general case, analytical solutions are not known), we know of no way to further characterize the set of sources for which the assumptions of Corollary 5.8 hold. However, we refer to [1, 8, 17], where for certain classes of analytically given convolution operators similar results to Corollary 5.8 can be guaranteed under simple structural assumptions on the source, such as, e.g., a minimum separation distance between the support points of (5.1). In our situation, we will investigate the assumptions of Corollary 5.8 numerically in section 7. The numerical results suggest that, even in the case of an arbitrary number of measurements, the source condition holds only in some cases. However, for a special choice of the weight, reconstruction of a single point source can be guaranteed.

5.1. **Exact reconstruction of a single source.** In this section we prove that, using the weight $w_{\Omega,2}$ as defined in (4.4), a source consisting of a single Dirac-delta function can always be reconstructed using the weighted problem. We first consider the noise free case:

Proposition 5.9. *Suppose that $w_{\Omega,2}^n(x) > 0$ for all $x \in \Omega_c$, $n = 1, \dots, N$. Let $u^\star = \boldsymbol{u}^\star \delta_{x^\star}$ with $x^\star \in \Omega_c \setminus \Xi$, $\boldsymbol{u}^\star \in \mathbb{C}^N$ and consider noise-free observations $p_d = Su^\star$. Then, for any $\alpha > 0$ and $w = w_{\Omega,2}$ the function*

$$\widehat{u} = \widehat{\boldsymbol{u}}\delta_{x^\star} \quad \textit{with} \quad \widehat{\boldsymbol{u}} = \max\{0,\ 1 - \alpha/|\boldsymbol{u}^\star w(x^\star)|_{\mathbb{C}^N}\}\boldsymbol{u}^\star$$

is a solution of (4.1). *Furthermore,* $u^\dagger = u^\star$ *solves the corresponding minimum norm problem defined as in section 5.*

Proof. We verify that the first order conditions from Theorem 4.9 are fulfilled. First, we compute $\widehat{\xi} = -S^*(S\widehat{u} - p_d)$ at every point and frequency. We directly obtain that

$$\widehat{\xi}_n(x) = \sum_{m=1}^{M} \bar{G}_n^{x_m}(x) G_n^{x_m}(x^\star)(\boldsymbol{u}_n^\star - \widehat{\boldsymbol{u}}_n), \quad x \in \Omega_c,\ n = 1, \ldots, N.$$

We compute that $(\boldsymbol{u}_n^\star - \widehat{\boldsymbol{u}}_n) = \min\{1,\ \alpha/|w(x^\star)\boldsymbol{u}^\star|_{\mathbb{C}^N}\}\boldsymbol{u}_n^\star$. Introducing the rescaled Green's functions $h_m^n = G_n^{x_m}/w^n$, we obtain

$$\widehat{\xi}_n(x)/w^n(x) = \sum_{m=1}^{M} \bar{h}_m^n(x) h_m^n(x^\star) \min\{1,\ \alpha/|w(x^\star)\boldsymbol{u}^\star|_{\mathbb{C}^N}\} w^n(x^\star)\boldsymbol{u}_n^\star. \tag{5.4}$$

By the definition of $w = w_{\Omega,2}$, we compute that $|h^n(x)|_{\mathbb{C}^M} = \sqrt{\sum_m |h_m^n(x)|^2} = |G^n(x)|_{\mathbb{C}^M}/w_{\Omega,2}(x) = 1$ for all $x \in \Omega_c \setminus \Xi$. Therefore, we can apply the Cauchy-Schwarz inequality to the term $\sum_m \bar{h}_m^n(x) h_m^n(x^\star)$ in (5.4) and obtain

$$|\widehat{\xi}_n(x)|/w^n(x) \le \min\{1,\ \alpha/|w(x^\star)\boldsymbol{u}^\star|_{\mathbb{C}^N}\} w^n(x^\star)|\boldsymbol{u}_n^\star|.$$

Summing the squares of both sides and taking the square root, we derive that

$$|\widehat{\xi}(x)/w(x)|_{\mathbb{C}^N} \le \min\{|w(x^\star)\boldsymbol{u}^\star|_{\mathbb{C}^N},\ \alpha\} \le \alpha, \quad x \in \Omega_c.$$

In the case that $\alpha < |w(x^\star)\boldsymbol{u}^\star|_{\mathbb{C}^N}$, it remains to verify the optimality condition for $\widehat{\boldsymbol{u}}$: Taking $x = x^\star$, we have $\sum_m \bar{h}_m^n(x^\star) h_m^n(x^\star) = 1$ in (5.4), and it follows that

$$\widehat{\xi}_n(x^\star)/w^n(x^\star) = \alpha w^n(x^\star)\boldsymbol{u}_n^\star/|w(x^\star)\boldsymbol{u}^\star|_{\mathbb{C}^N}, \quad n = 1, \ldots, N,$$

which implies the desired condition, since $\widehat{\boldsymbol{u}}$ and $\boldsymbol{u}^\star$ are scalar multiples of each other. Thus, $\widehat{u}_\alpha$ solves the weighted problem by Theorem 4.9.

In the case $\alpha = 0$, we show that the solution of the dual problem is given by $y^\dagger = p_d/|w(x^\star)\boldsymbol{u}^\star|_{\mathbb{C}^N}$. In light of Proposition 5.4, we have to verify that $\xi^\dagger = S^* y^\dagger$ fulfills the source condition, i.e., $\xi^\dagger/w \in \partial\|wu^\star\|_{\mathcal{M}(\Omega_c,\mathbb{C}^N)}$. We have

$$\begin{aligned} \xi_n^\dagger(x)/w^n(x) &= \sum_{m=1}^{M} \bar{h}_m^n(x) p_d^m/|w(x^\star)\boldsymbol{u}^\star|_{\mathbb{C}^N} \\ &= \sum_{m=1}^{M} \bar{h}_m^n(x) h_m^n(x^\star) w^n(x^\star)\boldsymbol{u}_n^\star/|w(x^\star)\boldsymbol{u}^\star|_{\mathbb{C}^N}. \end{aligned}$$

Similarly, it follows $\|\xi^\dagger/w\|_{\mathcal{C}(\Omega_c,\mathbb{C}^N)} \le 1$ and $\xi^\dagger(x^\star)/w(x^\star) = w(x^\star)\boldsymbol{u}^\star/|w(x^\star)\boldsymbol{u}^\star|_{\mathbb{C}^N}$, which implies the result by Proposition 5.3. □

Note that (5.9) also applies in the case of only one measurement, i.e. $M = 1$. In this case, for any $\xi = S^* y$ with $y \in \mathbb{C}^N$, the expression $\|\xi/w\|_{\mathbb{C}^N}$ is constant in the domain Ω, and any source $\boldsymbol{u}\delta_x$ for arbitrary $x \in \Omega_c \setminus \Xi$ and appropriate $\boldsymbol{u} \in \mathbb{C}^N$ solves the minimum norm problem. A criterion for $u^\star$ to be the unique solution, which can be derived by straightforward extension of the previous result, is given next.

Proposition 5.10. *In addition to the requirements of Proposition 5.9, assume that the observations for different source locations are complex linearly independent (i.e., there exist no $x, x' \in \Omega_c$, such that $S\delta_x = zS\delta_{x'}$ for $z \in \mathbb{C}$).*

Then the functions given in Proposition 5.9 are the unique solutions of the respective problems.

6. Optimization algorithm

We base the numerical optimization of (1.3) upon the successive peak insertion and thresholding algorithm proposed in [6]. It is based on iterates of the form $u^k = \sum_{j=1,\dots,N_d^k} \boldsymbol{u}_j^k \delta_{x_j^k}$ (with distinct x_j^k and $\boldsymbol{u}_j^k \neq 0$) and performs alternating steps, combining insertion of Dirac delta functions at new locations with removal steps.

For the convenience of the reader, we give a general description of the resulting procedure in Algorithm 1. Note, that the point insertion is performed at the maximum of the norm of the current adjoint state. For more details we refer to [6, Section 5]. The following convergence result is obtained there:

Algorithm 1 Successive peak insertion framework [6]

while "duality-gap large" **do**
1. Compute $\xi^k = S^*(Su^k - p_d)$. Determine $\hat{x}^k \in \arg\max_{x\in\Omega_c} |\xi^k(x)|_{\mathbb{C}^N}$.
2. Set $\theta^k = \begin{cases} 0, & \|\xi^k\|_{\mathcal{C}(\Omega_c,\mathbb{C}^N)} \leq \alpha, \\ -\left[\alpha^{-2}\|p_d\|^2/2\right] \xi^k(\hat{x}^k), & \text{else.} \end{cases}$
3. Select stepsize $s^k \in (0,1]$ and set $u^{k+1/2} = (1-s^k)u^k + s^k\theta^k\delta_{\hat{x}_k}$.
4. Set $\mathcal{A} = \operatorname{supp}(u^{k+1/2})$ and find $\boldsymbol{u}^{k+1} \in \mathbb{C}^{N\#\mathcal{A}}$ such that $u^{k+1} = U_{\mathcal{A}}(\boldsymbol{u}^{k+1})$ with $j(u^{k+1}) \leq j(u^{k+1/2})$.

Theorem 6.1 ([6, Theorem 5.8])**.** *Let the sequence u^k be generated by Algorithm 1. Then every subsequence of u^k has a weak-$*$ convergent subsequence that converges to a minimizer $\widehat{u}$. Furthermore:*

$$j(u^k) - j(\widehat{u}) \leq \frac{C}{k}.$$

To discuss different possible implementations of step 4 in Algorithm 1, we define for a ordered set of distinct points $\mathcal{A} = \{x_j \in \Omega_c \mid j = 1, \dots, \#\mathcal{A}\}$ the operator $U_{\mathcal{A}} \colon \mathbb{C}^{N\#\mathcal{A}} \to \mathcal{M}(\Omega_c, \mathbb{C}^N)$ by

$$U_{\mathcal{A}}(\boldsymbol{u}) = \sum_{j=1}^{\#\mathcal{A}} \boldsymbol{u}_j \delta_{x_j}.$$

The removal steps are based on the consideration of the finite-dimensional problem

$$\begin{aligned} \min_{\boldsymbol{u}\in\mathbb{C}^{NN_d}} j(U_{\mathcal{A}}(\boldsymbol{u})) &= \frac{1}{2}\|S(U_{\mathcal{A}}(\boldsymbol{u})) - p_d\|^2_{\mathbb{C}^{NK}} + \alpha\|U_{\mathcal{A}}(\boldsymbol{u})\|_{\mathcal{M}(\Omega_c,\mathbb{C}^N)} \\ &= \frac{1}{2}\|\boldsymbol{S}_{\mathcal{A}}\boldsymbol{u} - p_d\|^2_{\mathbb{C}^{NK}} + \alpha \sum_{j=1}^{\#\mathcal{A}} |\boldsymbol{u}_j|_{\mathbb{C}^N}, \end{aligned} \tag{6.1}$$

for $\mathcal{A}$ determined by an intermediate iterate and $(\boldsymbol{S}_{\mathcal{A}})_{j,n} = S\delta_{x_j} e_n$. Different concrete choices of step 4 are discussed in [6, Section 5]: it is suggested to perform one step of the well-know proximal gradient/iterative tresholding algorithm for the finite dimensional problem (6.1). In this way, step 5 is easy to implement, has a small cost (depending linearly on the current size of the support), and has the potential to set some coefficients to zero (by virtue of the soft shrinkage operator). Additional steps of the proximal gradient method could be performed, to possibly increase this "sparsifying" effect. Note that if we omit step 4, the size of the support will grow monotonically throughout the iterations due to the particular form of step 3 (except for the unlikely case that $s_k = 1$).

In our setting, we additionally know that solutions consisting of at most $2NM$ Dirac delta functions exist; see Corollary 3.6. Since the proof of the underlying result is constructive, it directly suggests an algorithm to remove excess point sources; see Proposition B.5.

Corollary 6.2. *For given $u^{k+1/2}$ with $\#\mathrm{supp}(u^{k+1/2}) > 2NM$, the algorithm from the proof of Proposition B.5 constructs a new iterate $u^{k+1} = U_{\mathcal{A}}(\boldsymbol{u})$, such that $\boldsymbol{u}_{\hat{\jmath}} = 0$ for one $\hat{\jmath}$ and $j(u^{k+1}) \leq j(u^{k+1/2})$.*

Proposition 6.3. *Suppose that step 5 of Algorithm 1 includes the procedure from Corollary 6.2 and that u^0 consists of at most $2NM$ Dirac delta functions. Then the iterates u^k and each weak-$*$ accumulation point $\widehat{u}$ of u^k consists of at most $2NM$ Dirac delta functions (in addition to the properties from Theorem 6.1).*

Proof. The bound on the support size for u^k is a direct consequence of Corollary 6.2. The bound for the limit follows from a general result on the weak-$*$ convergence of measures consisting of a uniformly bounded number of Dirac delta functions; see Appendix C. ☐

Additionally, [6] suggests acceleration strategies based on point moving and merging. Since they cannot be easily realized in our numerical setup using $\mathcal{C}^0$ finite elements (see section 7), we do not discuss them here. Alternatively, we suggest to solve the subproblem (6.1) exactly (up to machine precision) to accelerate the convergence. The resulting procedure is given in Algorithm 2. Since the

Algorithm 2 Primal-Dual-Active-Point strategy

while "duality-gap large" **do**
1. Calculate $\xi^k = S^*(Su^k - p_d)$. Determine $\hat{x}^k \in \arg\max_{x\in\Omega_c} |\xi^k(x)|_{\mathbb{C}^N}$.
2. Set $\mathcal{A} = \mathrm{supp}(u^k) \cup \{\hat{x}^k\}$, compute a solution $\widehat{\boldsymbol{u}} \in \mathbb{C}^{N\#\mathcal{A}}$ of (6.1) with $\#\,\mathrm{supp}(U_{\mathcal{A}}(\widehat{\boldsymbol{u}})) \leq 2NM$, and set $u^{k+1} = U_{\mathcal{A}}(\widehat{\boldsymbol{u}})$.

point insertion is the same in both algorithms, Algorithm 2 is a special case of Algorithm 1.

Proposition 6.4. *The iterates of Algorithm 2 coincide with the iterates of Algorithm 1, if in step 4, $\boldsymbol{u}^{k+1}$ is chosen as a solution $\widehat{\boldsymbol{u}} \in \mathbb{C}^{N\#\mathcal{A}}$ of (6.1).*

Proof. This is a direct consequence of the fact that step 1 and the choice of $\mathcal{A}$ coincide for both algorithms, and that $j(U_{\mathcal{A}}(\widehat{\boldsymbol{u}})) \leq j(u^{k+1/2}) \leq j(u^k)$; see [6, Proposition 5.6]. ☐

Remark 5. Another possible stopping criterion for Algorithm 2 would be the condition that the active set $\mathcal{A}$ coincides in two subsequent iterations k and $k+1$ i.e., that $\hat{x}^{k+1} \in \mathcal{A}(u^{k+1})$ in step $k+1$. Clearly, if this holds true, we have

$u^{k+1} = u^{k+2} = \widehat{u}$. In fact, the optimality of u^{k+1} can be obtained in this situation by formulating the optimality conditions of (6.1) from step k for $u^{k+1} = U_{\mathcal{A}}(\widehat{\boldsymbol{u}})$, concluding that $\hat{x}^{k+1} \in \mathcal{A}(u^{k+1})$ implies that $\|\xi^{k+1}\|_{\mathcal{C}(\Omega_c,\mathbb{C}^N)} \leq \alpha$ and verifying the first order conditions from Corollary 3.7, which are sufficient for optimality.

It remains to address the cost associated with the numerical solution of subproblem (6.1). It is well-known that this problem can be reformulated as a second order cone constrained linear optimization problem, by introducing $\#\mathcal{A}+1$ additional variables. Such problems can be solved efficiently by interior point methods. Since we can bound the number of active points $\#\mathcal{A}$ a priori by $2NM+1$, the cost for the approximate numerical solution of (6.1) (up to machine precision) can be regarded as a constant; see, e.g., [5]. In practice, we choose to implement a semismooth Newton method; see, e.g., [30]. While there are no complexity bounds for this class of methods, the local superlinear convergence properties (which, in contrast to interior point methods, allows for warm starts) makes this alternative seem appealing, since we have a potentially good initial guess for $\widehat{\boldsymbol{u}}$ from the previous iteration.

7. Numerical Results

In this section we briefly describe the discretization methods used for the solution of the Helmholtz equation in a bounded domain and for the sources from $\mathcal{M}(\Omega_c,\mathbb{C}^N)$. Let $p = p_n^1 + \mathrm{i}p_n^2$, $n = 1,\dots,N$ the solution of (2.1) for the control $u_n = u_n^1 + \mathrm{i}u_n^2$, $u \in \mathcal{M}(\Omega_c,\mathbb{C}^N)$. For the numerical computations we rewrite the state equation (2.1) in following equivalent real-valued form

$$\left\{ \begin{aligned} \begin{pmatrix} -\Delta - k_n^2 I & 0 \\ 0 & -\Delta - k_n^2 I \end{pmatrix} \begin{pmatrix} p_n^1 \\ p_n^2 \end{pmatrix} &= \frac{1}{w^n} \begin{pmatrix} u_n^1|_\Omega \\ u_n^2|_\Omega \end{pmatrix} \quad \text{in } \Omega, \\ \begin{pmatrix} \partial_\nu & \kappa_n \chi_{\Gamma_Z} \\ -\kappa_n \chi_{\Gamma_Z} & \partial_\nu \end{pmatrix} \begin{pmatrix} p_n^1 \\ p_n^2 \end{pmatrix} &= \frac{1}{w^n} \begin{pmatrix} u_n^1|_\Gamma \\ u_n^2|_\Gamma \end{pmatrix} \quad \text{on } \Gamma, \end{aligned} \right. \tag{7.1}$$

where w is one of the weight functions introduced in Section 4. Based on this formulation of the state equation we employ linear finite elements on a triangulation of Ω for the approximation of the state variables p_n^1 and p_n^2; cf. [2, 15, 26]. We only mention that the discretized state equation has unique and stable solutions (p_h^1, p_h^2) for a small enough grid size h; see, e.g., [2, Theorem 4.4].

We denote the set of grid nodes in the triangulation with $\mathcal{N}$. Moreover we denote the number of grid points with N_h and denote number of grid nodes in Ω_c with N_c. Corresponding to the discretization of the state space by finite elements, we discretize the control space by Dirac-delta functions in the gird nodes (see [10]):

$$\mathcal{M}_h = \left\{ u \in \mathcal{M}(\Omega_c,\mathbb{C}^N) \;\middle|\; u = \sum_{i=1}^{N_c} \boldsymbol{u}_i \delta_{x_i},\ \boldsymbol{u}_i \in \mathbb{C}^N, x_i \in \Omega_c \cap \mathcal{N} \right\}. \tag{7.2}$$

Since the measure is discretized in the grid nodes, we only need to compute the values of the weight w in the grid nodes to obtain a fully discrete problem. For instance, for the weight function $w^n_{\Omega,2} = \sqrt{\sum_{m=1}^M |G_n^{x_m}|^2}$, the functions $G_n^{x_m}$ are approximated again by linear finite elements. Based on the pointwise values of the finite element approximations we obtain a discrete approximation of the given weight in the grid nodes.

We introduce the discrete reweighed observation mapping $S_h^w\colon \mathcal{M}_h \to \mathbb{C}^{NM}$ defined by

$$S_h^w\colon u \mapsto \{p_{n,h}^1(x_m) + \mathrm{i}p_{n,h}^2(x_m)\}_{n,m=1}^{N,M}.$$

Based on the operator S_h^w we formulate the reweighed discrete control problem

$$\min_{u\in\mathcal{M}_h} j_h(u) = \frac{1}{2}\sum_{m=1}^{M} |(S_h^w u)_m - p_d^m|_{\mathbb{C}^N}^2 + \alpha\|u\|_{\mathcal{M}(\Omega_c,\mathbb{C}^N)}. \tag{7.3}$$

For an $u \in \mathcal{M}_h$ the regularization functional has the form

$$\alpha\|u\|_{\mathcal{M}(\Omega_c,\mathbb{C}^N)} = \alpha\sum_{j=1}^{N_c} |\boldsymbol{u}_j|_{\mathbb{C}^N}.$$

Thus, problem (7.3) is a finite dimensional non-smooth and convex optimization problem. There are several algorithms which can be used for its solution. For example, the CVX toolbox [21] reformulates the problem as a cone constrained problem and solves the resulting problem using an interior point method. While highly efficient for medium sized problems, the performance of such a method suffers dramatically from the high dimension $2NN_c$ of the optimization variable in problem (7.3) (in the case of a fine discretization).

Finally, we implement the algorithms from section 6 on the discrete level. To adapt Algorithm 1 and Algorithm 2 to the discrete level, it suffices to note that the maximization of the adjoint variable ξ^k needs to be performed only over the grid points, which is done by a direct search. The other steps can be implemented directly. Since the dimension of the observation $2NM$ is low in comparison to $\dim \mathcal{M}_h = 2NN_c$, we build up the matrix representation $(\boldsymbol{S}^w)^* \in \mathbb{C}^{NM\times NN_c}$ of $(S_h^w)^*$ in a preprocessing step. This step involves M-times the solution of the discrete adjoint state equation. By transposition we get the matrix representation $\boldsymbol{S}^w$ of S_h^w. Note that this matrix is often referred to as the mixing matrix of a microphone array in Beamforming applications; see [32]. Thus, the evaluation of the solution operator and the adjoint equation needed for the application of Algorithm 1 resp. 2 reduces to a matrix vector multiplication. Due to the convergence analysis on the continuous level, we can expect the algorithms to behave independently of the number of grid points, where the cost of each iteration scales linearly in N_c.

7.1. **Interpretation of discrete solutions.** It is known that a discretization of a measure on a finite grid introduces artifacts: Roughly speaking, a source present in the continuous problem at a off-grid location tends to appear spread out over the adjacent grid cells, which artificially increases the number of support points in the discrete solution, and makes the direct interpretation of the numerical solutions difficult. For a theoretical analysis of this effect we refer to [17]. For practical purposes, we employ the following post-processing strategy: First, we build the connectivity graph of the sparsity pattern of the finite element discretization, and interpret all point sources less than two nodes away from each other as part of a cluster. Then, for each cluster we replace the sources of hat cluster by a source located at the center of gravity of the cluster with a coefficient given by the sum of the coefficients. Mathematically, this can be regarded as an interpolation operation on the space of measures, which introduces an additional error proportional to h under reasonable assumptions.

7.2. **Numerical experiments.** In this section we conduct several numerical experiments based on an acoustic inverse source problem involving the Helmholtz equation. In all considered scenarios we are given a computational domain Ω with reflecting as well as absorbing boundary conditions. We give examples to demonstrate the applicability of the general approach, and investigate the influence of the choice of the weight w and the performance of the presented algorithms. In all examples, we use the following setting:

- The computational domain is given by a square of four by four meters, i.e., $\Omega = [0, 4]^2$.
- The computational grid $\mathcal{T}_h$ is given by an uniform triangular discretization of Ω with $h = \sqrt{2}/2^l$ with grid level $l \in \{6, \dots, 9\}$.
- Two reflecting walls Γ_N are located on the left and top and two absorbing walls Γ_Z (with $\kappa_n = k_n$) on the bottom and right.
- The speed of sound is set to $c = 345\,[\mathrm{m/s}]$.

7.2.1. *Deterministic comparison of weights.* The results of Proposition 5.9 show that one point source can be exactly recovered in the noise free case for the weighted approach (4.1). However, we can construct a simple example, which numerically demonstrates that the reconstruction based on the non-weighted approach (3.1) does not necessarily yield the exact positions and intensities in this scenario. To this purpose, we choose an exact source located close to the reflecting boundaries of Ω and compute a minimum norm solution for different problem formulations. More precisely, we set $u^\star = e^{i\pi/4}\delta_{x^\star}$ with $x^\star = (0.5; 3.75)$. Furthermore, for simplicity, we consider the case with only one frequency $\omega = 2\pi\,261.6$, which corresponds to the tone C4, and three microphones located in $(3.75, 1)$, $(3.75, 2)$, $(3.75, 3)$ as depicted in Figure 1.

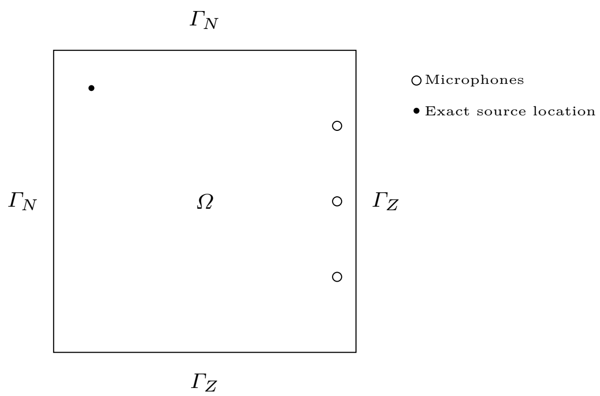
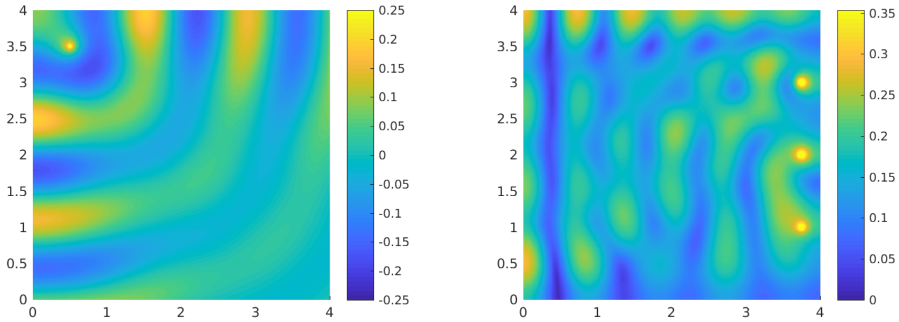


FIGURE 1. The computational domain Ω, the array of microphones and the exact source position.

Since we compare different problems settings under ideal conditions, we consider noise-free observations which are generated on the same grid as the subsequent computations. Therefore, we set $p_d = p_h(u^\star)$ generated by solving the discrete Helmholtz equation (3.2) with the exact source $u^\star$. In Figure 2a the real part of the acoustic pressure $p(u^\star)$ is displayed. Circular waves are generated from the point source and intensified by the reflections on Γ_N. Figure 2b shows the



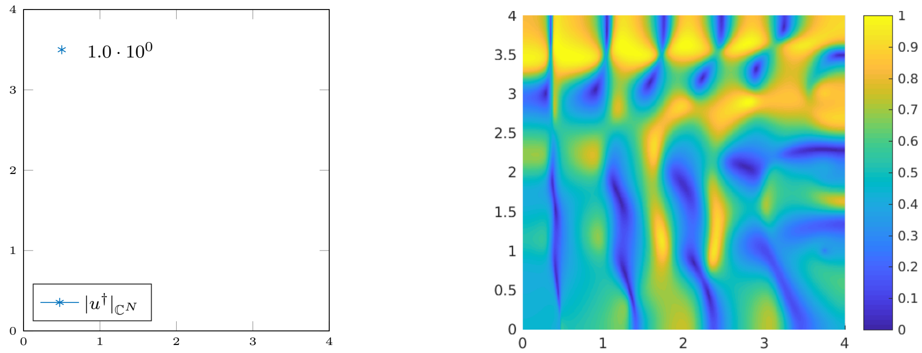
(a) Real part of the acoustic pressure $p_h(u^\star)$.

(b) Discrete approximation of $w_{\Omega,2}$.

Figure 2. Exact pressure and weight $w = w_{\Omega,2}$.

weight $w = w_{\Omega,2} = \sqrt{\sum_{m=1}^{M} |G^{x_m}|^2}$. As mentioned before, the value of the weight at point in the domain corresponds to the magnitude of the signal that will be received at the microphones. We clearly see that w has a relatively low value in a neighborhood of the exact source position. This behavior of w is caused by negative interference of the generated and reflected waves. Furthermore, we clearly observe the large values of the weight close to the microphones.

In the following, we numerically approximate the minimum norm solutions $u^\dagger$ for different weights. To this purpose, we solve the respective discrete problems for a decreasing sequence of cost parameters ($\alpha = 10^{-0}, \ldots, 10^{-10}$) up to machine precision (using Algorithm 2). Then, we take the solution $\widehat{u}_\alpha$ for the smallest α as an approximation of $u^\dagger$ (which is justified by Corollary 5.8). Furthermore, an approximation of the element $\xi^\dagger$ from the source condition (5.3) is given by $\widehat{\xi}_\alpha = -S^*(S\widehat{u}_\alpha - p_d)/\alpha$.

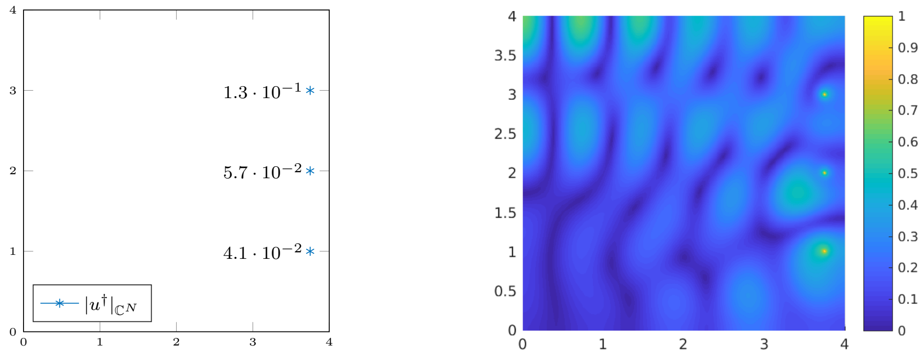


(a) Positions and source intensities of the minimum norm solution $u^\dagger$.

(b) Discrete approximation of $|\xi^\dagger(x)/w|_{\mathbb{C}^N}$.

Figure 3. Noise free reconstructions for $\Omega_c = \Omega$, weight $w = w_{\Omega,2}$.

We give the results for $w = w_{\Omega,2}$ in Figure 3a. Here, for the reconstruction we admit all possible sources and set $\Omega_c = \Omega$. In agreement with Proposition 5.9 we observe that the support of the solution is recovered exactly, and that the coefficient coincides to the exact one up to the seventh digit. Moreover, a close inspection of the variable $|\xi^\dagger(x)/w|_{\mathbb{C}^N}$ shows that its maximum value one is uniquely attained at the exact source position; the next biggest local minimum has a value of ~ 0.995. This demonstrates uniqueness of the discrete minimum norm solution in this case (cf. Proposition 5.7).

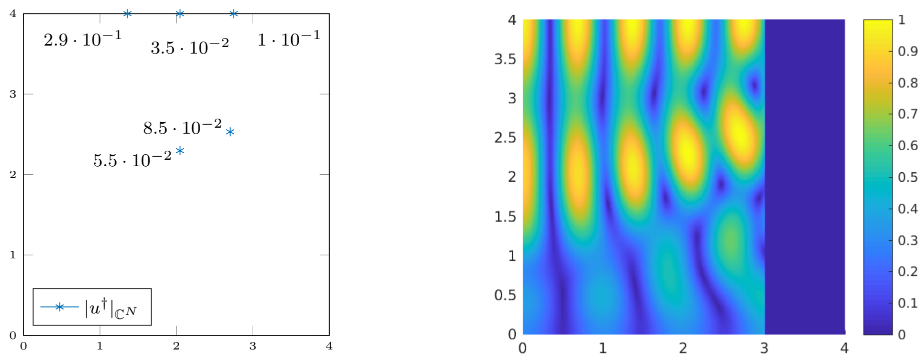


(a) Positions and source intensities of the minimum norm solution $u^\dagger$.

(b) Discrete approximation of $|\xi^\dagger(x)|_{\mathbb{C}^N}$.

FIGURE 4. Noise free reconstructions for $\Omega_c = \Omega$, no weight ($w \equiv 1$).

Next, we consider the case without weight. According to Proposition 4.1 the corresponding problem with $\Omega_c = \Omega \backslash \Xi$ has no solution since there exists vanishing sequences of point sources which generate the exact measurements and converge to the positions of the microphones. However, in the discrete setting the problem always has a solution, since the discrete Green's functions are bounded by a mesh-dependent constant. We give the numerical results in Figure 4a. Here, the minimum norm solution $u^\dagger$ consists of three point sources located in the microphone positions. The maximum of the absolute value of the adjoint state is assumed only there; see Figure 4b. Note that this numerical solution is highly sensitive to the grid resolution. In fact, for $h \to 0$ the minimum norm solution and dual variable converge to zero.



(a) Positions and source intensities of the minimum norm solution $u^\dagger$.

(b) Discrete approximation of $|\xi^\dagger(x)|_{\mathbb{C}^N}$.

FIGURE 5. Results for $\Omega_c = [0, 3] \times [0, 4]$, no weight ($w \equiv 1$).

To obtain a well-posed optimization problem without weight we choose the control domain $\Omega_c = [0, 3] \times [0, 4]$, which excludes the observation positions. The results are given in Figure 5a, where we observe that the optimal solution consists of five point sources: three are located on the reflecting boundary Γ_N and three are located in the interior of the domain. The corresponding function $|\xi^\dagger(x)|_{\mathbb{C}^N}$ attains its global maximum on the support points of $u^\dagger$. However, the region close to the exact source position assumes a visibly lower function value, and no source is placed there. This can be connected to the negative interference at this point; cf. Figure 2b.

These examples show that even in simple settings the reconstruction results of the non-weighted approach (3.1) is affected by negative interference caused by the reflecting boundaries, as well as the fact that the adjoint state takes arbitrarily large values close to the microphone positions.

7.2.2. *Statistical comparison of weights.* Now, we consider a more involved problem setup to evaluate the reconstruction quality for different weights. We consider the same model as before, but consider the frequencies $\omega = 2\pi(349.2, 523.3, 659.3)$ (corresponding to F4, C5, and E5). The number of microphones is increased to 30, and the control domain is chosen as $\Omega_c = [0,3] \times [0,4]$, which does not contain the microphone locations at (x_1, x_2) with $x_1 = 3.25$ and $x_1 = 3.75$ and x_2 regularly spaced from 0 to 4; see Figure 7a. All computation are performed on grid level $l = 8$.

To evaluate to reconstruction quality of different weights, we follow a statistical approach: for each number of point sources $N_d^\star \in \{1, 2, \dots, 5\}$, we generate a random source by selecting $N_d^\star$ random indices from the mesh nodes on the control domain and generating corresponding random coefficients by drawing from a multivariate complex Gaussian distribution with unit variance. Then, we compute a minimum norm solution $(P_{0,w})$ from the corresponding exact observations for the given weight, which is either $w \equiv 1$ or $w = w_{\Omega,2}$. Here, we again approximate the minimum norm solution by the solution for a value of $\alpha = 10^{-9}$, which we compute by a continuation strategy in the regularization parameter using Algorithm 2.

Finally, we evaluate the average reconstruction error for each weight. Since the generalized Bregman distance is multivalued, we focus on two simple citeria. The first is simply the relative difference of the norms with respect to the employed weight,

$$e_1 = \left[\|u^\star\|_{\mathcal{M}_w(\Omega_c,\mathbb{C}^N)} - \|u^\dagger\|_{\mathcal{M}_w(\Omega_c,\mathbb{C}^N)}\right] / \|u^\star\|_{\mathcal{M}_w(\Omega_c,\mathbb{C}^N)} \tag{7.4}$$

Note that it can be easily verified that $\|u^\star\|_{\mathcal{M}_w(\Omega_c,\mathbb{C}^N)} - \|u^\dagger\|_{\mathcal{M}_w(\Omega_c,\mathbb{C}^N)} \in D(u^\star, u^\dagger)$ (for the specific choice $\xi = S^* y^\dagger$), which relates this criterion to the Bregman distance; cf. Theorem 5.6. The results are given in Figure 6a. We observe that the difference is smaller for the weight $w_{\Omega,2}$, and that it is zero for the case of one source, as predicted by theory. However, we can expect the norm difference to severely underestimate the reconstruction error. Moreover, the results for different weights are not directly comparable, due to the fact that the error criterion itself depends on the weight. Therefore, we also consider a second error criterion, which is based on convolution. We introduce the componentwise convolution operator $S^\sigma_{\text{heat}} \colon \mathcal{M}(\Omega, \mathbb{C}^N) \to L^1(\Omega, \mathbb{C}^N)$, which computes the solution at time $T = \sigma^2/2$ of the heat equation (endowed with homogeneous Neumann boundary conditions on the domain Ω) with the given initial data at time zero. Then we define the second error criterion by

$$e_2 = \|S^\sigma_{\text{heat}}(u^\star - u^\dagger)\|_{L^1(\Omega,\mathbb{C}^N)}/\|u^\star\|_{\mathcal{M}(\Omega,\mathbb{C}^N)}. \tag{7.5}$$

Here, we compare the reconstruction error in the canonical norm after convolution with a regular kernel with approximate width σ. Roughly speaking, we can expect small errors in the source location to lead to small error terms (which is not the case if we apply the total variation norm directly), whereas location errors larger than σ lead to big error contributions. Mathematically, the backwards uniqueness property of the heat equation guarantees that $e_2 = 0$ can only occur for $u^\dagger = u^\star$. We implement S^σ_{heat} by a finite element approximation on the given grid and an implicit Euler time discretization (with five steps). The results for $\sigma = 0.2$ and $\sigma = 0.05$ are given in Figures 6b and 6c, respectively. We observe that, although the errors increase for more strict error criteria, the average errors are consistently smaller when the weight $w_{\Omega,2}$ is employed.

N_d^*	$w \equiv 1$	$w = w_{\Omega,2}$
1	0.0087	0.0000
2	0.0233	0.0030
3	0.0599	0.0174
4	0.0867	0.0404
5	0.1443	0.0754

(a) Average relative norm error (7.4).

N_d^*	$w \equiv 1$	$w = w_{\Omega,2}$
1	0.0875	0.0000
2	0.1894	0.0387
3	0.4364	0.2042
4	0.6416	0.4394
5	0.8326	0.6691

(b) Convolution error (7.5) with $\sigma = 0.2$.

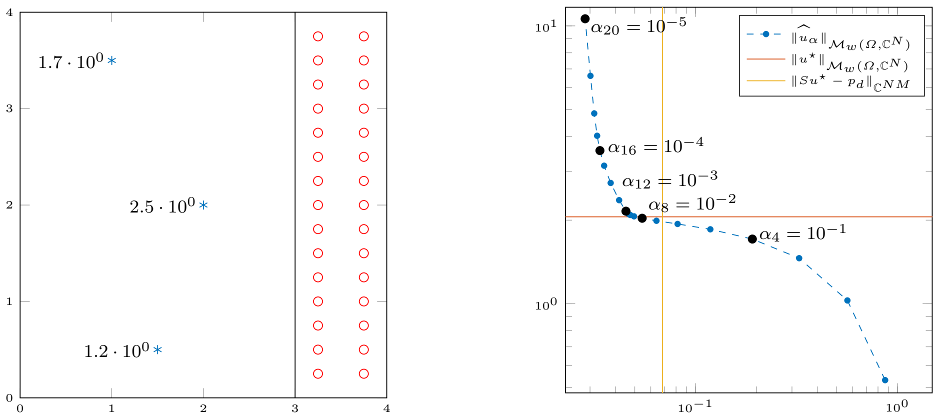
N_d^*	$w \equiv 1$	$w = w_{\Omega,2}$
1	0.1453	0.0000
2	0.2660	0.0625
3	0.6129	0.2971
4	0.8689	0.6271
5	1.1181	0.9556

(c) Convolution error (7.5) with $\sigma = 0.05$.

FIGURE 6. Average reconstruction error for 200 randomly generated sources with different numbers of point sources $N_d^\star$.

7.2.3. *Comparison of algorithms.* Now, we evaluate the practical performance of the algorithms from section 6. We consider the same setting as in the previous section (frequencies $\omega = 2\pi(349.2, 523.3, 659.3)$ and 30 microphones). We recover a source consisting of three point sources as depicted in Figure 7a with random coefficients (drawn from a multivariate complex Gaussian distribution with unit variance). The control domain is chosen as $\Omega_c = [0, 3] \times [0, 4]$ and the weight $w_{\Omega,2}$ is employed in all experiments.

We want to study the algorithms for a setting with noise and useful values of the parameter α. Therefore, we compute synthetic measurements on the finest grid level $l = 9$ and perturb them by additive Gaussian noise, such that $\|Su^\star - p_d\|/\|Su^\star\| = 5\%$. We then solve the problem on a coarser grid level $l = 8$, to also take into account a possible discretization error. To determine a useful range of regularization parameters, we numerically compute an L-curve: we solve the problem $(P_{\alpha,w})$ for a sequence of regularization parameters $\alpha_j = 10^{-j/4}$, $j = 0, 1, \ldots, 20$ and plot the norm of the solution $\widehat{u}_\alpha$ over the data misfit term $\|S\widehat{u}_\alpha - p_d\|_{\mathbb{C}^{NM}}$; see Figure 7b. We observe that the data misfit term is reduced below the noise level at $\alpha_7 \approx 1.8 \cdot 10^{-2}$ (corresponding to the popular Morozov-criterion for the selection of a regularization parameter), and at $\alpha_9 \approx 5.6 \cdot 10^{-3}$ the norm of the reconstruction starts to exceed the norm of the exact solution $u^\star$. We conclude that practically relevant values of α are around 10^{-2} in this particular instance.

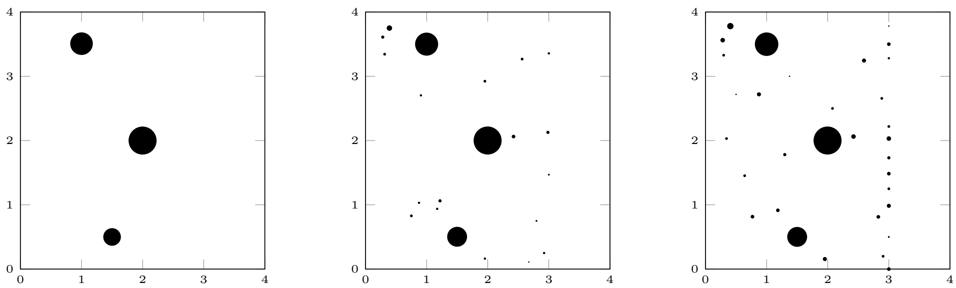


(a) Exact source locations and source intensities on the left, microphone locations on the right.

(b) Norms of the solutions $\widehat{u}_\alpha$ over the data misfit $\|S\widehat{u}_\alpha - p_d\|_{\mathbb{C}^{NM}}$ for different α for noisy observations p_d (5% noise).

FIGURE 7. Problem setup and L-curve at grid level $l = 8$.

In a first test, we compute reconstructions (on grid level $l = 8$) starting from an initial guess of $u^0 = 0$ for $\alpha = 10^{-1}, 10^{-2}, 10^{-3}$ with different algorithms. A visualization of the corresponding numerical solutions (computed with Algorithm 2 up to machine precision) is given in Figure 8.

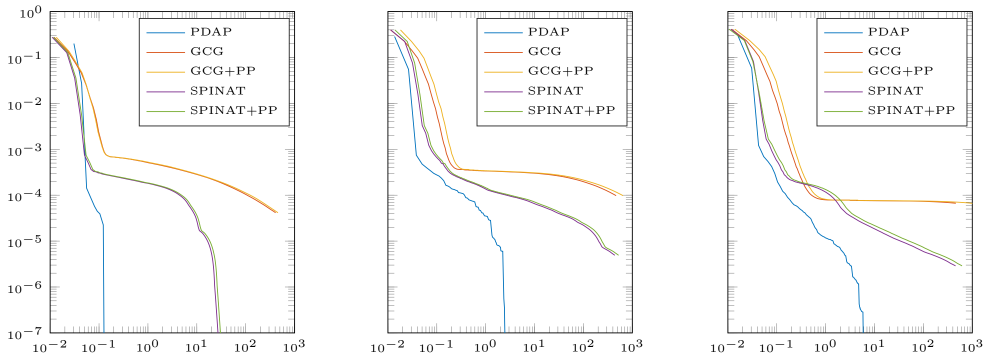


(a) $\alpha = 10^{-1}$ (b) $\alpha = 10^{-2}$ (c) $\alpha = 10^{-3}$

FIGURE 8. Visualization of the numerical reconstructions; each dot is one support point and dot area is proportional to source magnitude.

In the following, we consider Algorithm 2 (denoted by PDAP), and different versions of the accelerated conditional gradient method 1 without exact resolution of the subproblems. The unaccelerated version is denoted by GCG, and the version performing one iterative tresholding step for the subproblem in each iteration is denoted by SPINAT (cf. [6]). An suffix +PP denotes an additional application of the sparsifying post-processing step from Corollary 6.2. The numerical results are given in Figure 9, where we plot the evolution of the residual over the computation time (in seconds). We opt for computation times over the step counter k to account for the fact that one step of an accelerated method may be more costly. We note that all algorithms are implemented in MATLAB (version R2017a) and the computations are performed on a compute node with a Intel® Xeon® CPU

E5-2670 with eight cores at 2.60GHz. We observe that PDAP outperforms the

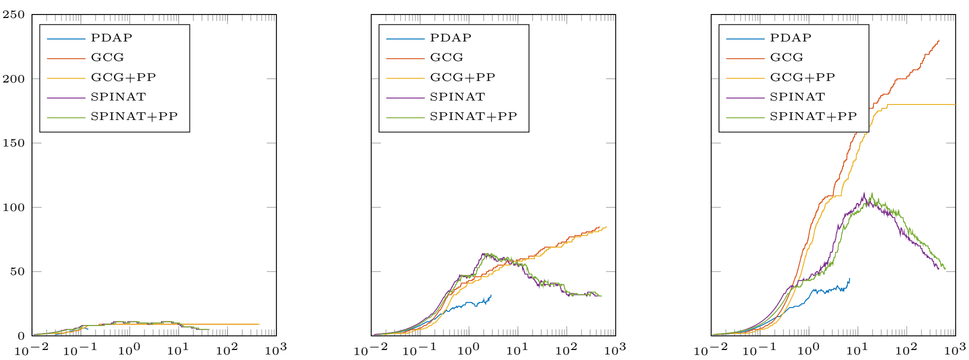


(a) $\alpha = 10^{-1}$ (b) $\alpha = 10^{-2}$ (c) $\alpha = 10^{-3}$

FIGURE 9. Residuals $j(u^k) - j(\widehat{u}_\alpha)$ over computation time in s. for different α.

other versions in almost all situations. With the exception of $\alpha = 10^{-1}$ it is the only implementation that is able to solve the problem up the tolerance within the computational budget of 50000 iterations (in fact it performs 10, 96, and 129 iterations, respectively). We also see that SPINAT improves upon GCG, but not by as much as PDAP.

Additionally, we also give the current support size in Figure 10. In the case of



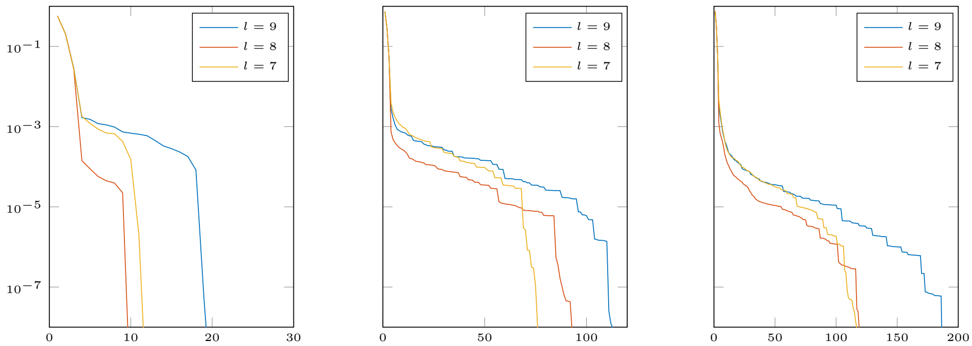
(a) $\alpha = 10^{-1}$ (b) $\alpha = 10^{-2}$ (c) $\alpha = 10^{-3}$

FIGURE 10. Support size $\#\operatorname{supp} u^k$ over computation time in s. for different α.

$\alpha = 10^{-1}$, which is under-fitting the data, all algorithms quickly identify a set of grid points which contain the support of the discrete numerical solution and thus effectively stop to insert new points. However, note that this is only the possible due to the finite grid, which limits the number of support point a priori. Note also that PDAP terminates once all support points have been identified; cf. Remark 5. In the other cases, the size of the support of the iterates is negatively impacted by the spurious point sources introduced from over-fitting the data. We note that for PDAP the support size of the iterates stays bounded by the numerical support of the optimal solution (see Figure 10), which keeps the cost of resolution of the subproblems small. The theoretical upper bound on the support size is $2NM = 180$, which is very pessimistic for this example, and only provides an advantage for GCG in the third setting.

Finally, we comment on the computation of the L-curve: Due to the fact that the solution for a big α can be used as an initial guess for a smaller α, the computation of the L-curve up to $\alpha_{12} = 10^{-3}$ with PDAP up to machine precision is not much more expensive than computing just the solution for the last α starting from zero. For instance, in this case the number of iterations for each α are $(1, 3, 3, 6, 2, 3, 7, 20, 27, 40, 34, 33, 49)$, which results in a combined ~ 24 seconds of computation time versus 129 iterations in ~ 7 seconds for just the last value.

7.2.4. *Mesh independence.* Additionally, we investigate the behavior of the algorithms with respect to the mesh width. Here, we only focus on PDAP, since we want to investigate if the improved convergence observed before depends on the finite discretization. Here, we compare iteration numbers, since the computation times are dominated by the assembly of the gradients p_h^k, which scales linearly in N_h. We give the results for the previous example on mesh levels $l = 7, 8, 9$ in Figure 11. We observe that although the number of iterations to reach machine precision increases on finer meshes, the functional residual follows a similar trajectory in the initial iterations. In the later iterations, the finite termination of the method is reached earlier on coarse grids. Concerning the maximal support of



(a) $\alpha = 10^{-1}$ (b) $\alpha = 10^{-2}$ (c) $\alpha = 10^{-3}$

FIGURE 11. Function residuals $j(u^k) - j(\widehat{u}_\alpha)$ over iterations for different mesh levels l.

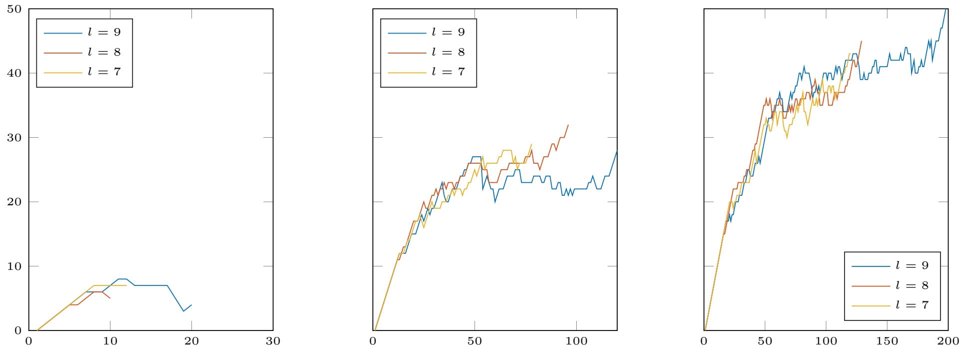
the numerical solution throughout the iterations, we observe that it seems to be dependent on α, but bounded by a similar constant independent of the grid level.

APPENDIX A. SPARSE MINIMIZATION WITH FINITE RANK OPERATORS

Let H_1 be a separable real Hilbert space, and $\mathcal{M}(D, H_1) = \mathcal{C}_0(D, H_1)^*$ be the associated space of vector measures. Introduce the solution operator

$$S \colon \mathcal{M}(D, H_1) \to H_2,$$

where H_2 is another separable real Hilbert space. S is assumed to be linear, and weak-$*$ to weak continuous (the weak-$*$ topology on the dual of the separable space



(a) $\alpha = 10^{-1}$ (b) $\alpha = 10^{-2}$ (c) $\alpha = 10^{-3}$

FIGURE 12. Support size $\#\operatorname{supp} u^k$ over iterations for different mesh levels l.

$\mathcal{C}_0(D, H_1)$ can be normed, therefore, this is the same as the sequential equivalent). Moreover, S can be written as the Banach space dual of a continuous operator

$$S^*\colon H_2 \to \mathcal{C}_0(D, H_1).$$

In this section, we give some results for the two abstract minimization problems relevant for this paper. Most of these results are slight generalizations of known results, which we could not directly find in the literature. We consider the problem

$$\min_{u \in \mathcal{M}(D,H_1)} \left[\frac{1}{2\alpha}\|Su - p_d\|_{H_2}^2 + \|u\|_{\mathcal{M}(D,H_1)}\right], \tag{P_α}$$

for given $p_d \in H_2$ and $\alpha > 0$. Note, that in contrast to (1.3), we have multiplied the objective function by $1/\alpha$, which obviously does not change the solution set, but leads to a more convenient form of the dual problem below. Moreover, we consider the associated minimum norm problem

$$\min_{u \in \mathcal{M}(D,H_1)} \|u\|_{\mathcal{M}(D,H_1)} \quad \text{subject to } Su = p_d, \tag{P_0}$$

for some $p_d = Su^\star$, $u^\star \in \mathcal{M}(D, H_1)$. It is know that under the general assumptions on S, both problems have solutions. This can be verified with the direct method of the calculus of variations. Moreover, the dual problem of (P_α),

$$\max_{y \in H_2} \left[(p_d, y)_{H_2} - \frac{\alpha}{2}\|y\|_{H_2}^2\right] \quad \text{subject to } \|S^* y\|_{\mathcal{C}_0(D,H_1)} \le 1, \tag{D_α}$$

has a unique solution, and the strong duality $\max\,(D_\alpha) = \min\,(P_\alpha)$ holds; see [6, Proposition 3.5] (the proof is only given for $H_1 = \mathbb{R}^n$, but works unmodified in the general case). For (P_0), the dual problem is given by

$$\max_{y \in H_2} \,(p_d, y) \quad \text{subject to } \|S^* y\|_{\mathcal{C}_0(D,H_1)} \le 1. \tag{D_0}$$

Since $p_d = Su^\star$, strong duality holds with $\sup\,(D_0) = \min\,(P_0)$; see [17, Proposition 13] (the proof is only given for $H_1 = \mathbb{R}$ and D equal to the torus, but works unmodified in the general setting).

Proposition A.1 ([17, Proposition 13]). *Let $p_d = Su^\star$, $u^\star \in \mathcal{M}(D, H_1)$. Then, strong duality (see, e.g., [18, Chapter 3.4]) holds for the problem (P_0) and the dual problem (D_0). If the dual problem admits a solution, any pair of solutions*

$(u^\dagger, y^\dagger)$ *to both problems is characterized by the subdifferential inclusion* $S^*y^\dagger \in \partial\|u^\dagger\|_{\mathcal{M}(D,H_1)}$.

In general, (D_0) does not necessarily have a solution. However, if S is a finite rank operator (the range of S or S^* is finite dimensional), the dual problem (D_0) admits a solution. This result is mentioned and used in [17]; however, since no proof is given, we provide one for the general setting above.

Proposition A.2. *Suppose that* $\operatorname{Ran} S$ *is finite dimensional and* $p_d \in \operatorname{Ran} S$. *Then, the dual problem* (D_0) *admits a solution. Suppose additionally that the adjoint* $S^*\colon H_2 \to \mathcal{C}_0(D, H_1)$ *is injective. Then, the above result holds for any* $p_d \in H_2$ *and the solution set of* (D_0) *is bounded.*

Proof. We first assume that $S^*\colon H_2 \to \mathcal{C}_0(D, H_1)$ is injective. Note that this implies H_2 is finite dimensional. In this case, (D_0) can be reformulated as a semi-infinite optimization problem, and the result can be deduced as an application of the general result [3, Theorem 5.99] (injectivity of S^* is equivalent to the regularity condition mentioned there). However, in our case, it can be also shown directly. In fact, any maximizing sequence for (D_0) is bounded: Take by contradiction $\{y_k\}$ with $\|S^*y_k\|_{\mathcal{C}_0(D,H_1)} \leq 1$ and $\|y_k\|_{H_2} \to \infty$. Considering the renormed sequence $\{\tilde{y}_k\}_{k\in\mathbb{N}}$ with $\tilde{y}_k = y_k/\|y_k\|_{H_2}$ there exists a subsequence denoted by the same symbol and a $\hat{y} \in H_2$ with $\tilde{y}_k \to \hat{y}$ and $\|\hat{y}\|_{H_2} = 1$ (since H_2 is finite dimensional). Consequently there holds

$$\|y_k\|_{H_2}\|S^*\tilde{y}_k\|_{\mathcal{C}_0(D,H_1)} = \|S^*y_k\|_{\mathcal{C}_0(D,H_1)} \leq 1.$$

From this we directly conclude that $\|S^*\hat{y}\|_{\mathcal{C}_0(D,H_1)} = 0$ since S^* is bounded. Then the injectivity of S^* implies a contradiction to $\|\hat{y}\|_{H_2} = 1$. Consequently, any minimizing sequence is bounded, and by using the continuity of S^*, it follows that there exits at least one optimal solution to (D_0). Boundedness of the solution set follows in the same way.

Now, we address the general case, where S^* is not necessarily injective, and show that it can be reduced to the previous case. Consider the problem

$$\max_{y\in\operatorname{Ran} S} (y, p_d) \quad \text{subject to } \|S^*y\|_{\mathcal{C}_0(D,H_1)} \leq 1. \tag{A.1}$$

Since $\operatorname{Ran} S$ is finite dimensional (and therefore a closed subspace), we have $(\operatorname{Ran} S)^\perp = \operatorname{Ker} S^*$, and $H_2 = \operatorname{Ran} S \oplus \operatorname{Ker} S^*$. For any $y \in H_2$ we have $y = y_1 + y_0$ with $y_1 \in \operatorname{Ran} S = (\operatorname{Ker} S^*)^\perp$ and $y_0 \in \operatorname{Ker} S^*$. Let $u^\star \in \mathcal{M}(D, H_1)$ be an element with $Su^\star = p_d$ which exists according to our assumptions. Then we have

$$(p_d, y) = \langle u^\star, S^*y_1\rangle = (p_d, y_1), \quad \text{and } \|S^*y\|_{\mathcal{C}_0(D,H_1)} = \|S^*y_1\|_{\mathcal{C}_0(D,H_1)}$$

which implies that (A.1) and (D_0) have the same value. Moreover, the restricted operator $S^*|_{\operatorname{Ran} S}\colon \operatorname{Ran} S \to \mathcal{C}_0(D, H_1)$ is injective. Using the result from before, (A.1) admits a solution, and for any solution y_1 and any $y_0 \in \operatorname{Ker} S^*$, $y = y_1 + y_0$ is a solution of (D_0). □

Appendix B. Extremal solutions

Since the dual problems (D_α) and (D_0) fall into the category of *semi-infinite* optimization problems, it follows that solutions of (P_α) and (P_0) consisting of finitely many Dirac delta functions exist; see, e.g., [3, Section 5.4.2].

For the convenience of the reader, we provide a direct proof, which also leads to an algorithmic strategy for reducing the support of any suboptimal point of (P_α) or (P_0). To this purpose, we analyze the corresponding solution sets, which we denote for $\alpha \geq 0$ by

$$U_{p_d,\alpha} = \{u \in \mathcal{M}(D, H_1) \mid u \text{ solves } (P_\alpha) \text{ for } \alpha > 0 \text{ or } (P_0) \text{ for } \alpha = 0\}.$$

This is a convex bounded subset of $\mathcal{M}(D, H_1)$. Furthermore the following properties are easily derived.

Proposition B.1. *Let $\widehat{u} \in U_{p_d,\alpha}$ be arbitrary, and $\widehat{p} = S\widehat{u}$. For all elements $u \in U_{p_d,\alpha}$ we have*

$$Su = \widehat{p}, \qquad \|u\|_{\mathcal{M}(D,H_1)} = \|\widehat{u}\|_{\mathcal{M}(D,H_1)}.$$

Proof. The statement is clear for $\alpha = 0$, where $\widehat{p} = p_d$. For $\alpha > 0$ the first part follows from the strict convexity of the tracking term and the linearity of S. Therefore, the value of the first term of the objective assumes a unique value for all optimal solutions. By the optimality follows that also the second term must be of the same value for all optimal solutions. □

As a corollary, we obtain a characterization of $U_{p_d,\alpha}$.

Corollary B.2. *Let $\widehat{u} \in U_{p_d,\alpha}$ be arbitrary, and $\widehat{p} = S\widehat{u}$. It holds,*

$$\begin{aligned} U_{p_d,\alpha} &= \{u \in \mathcal{M}(D, H_1) \mid Su = \widehat{p} \text{ and } \|u\|_{\mathcal{M}(D,H_1)} = \|\widehat{u}\|_{\mathcal{M}(D,H_1)}\} \\ &= \{u \in \mathcal{M}(D, H_1) \mid Su = \widehat{p} \text{ and } \|u\|_{\mathcal{M}(D,H_1)} \leq \|\widehat{u}\|_{\mathcal{M}(D,H_1)}\}. \end{aligned}$$

Now, we recall the concept of extremal points of convex set: A point in the convex set $U_{p_d,\alpha}$ is called *extremal*, if it can not be written as a nontrivial convex combination of other elements of $U_{p_d,\alpha}$. Furthermore, we have the theorem of Krein and Milman.

Proposition B.3. *The closure (in the sense of the weak-∗ topology) of the convex combinations of the extremal points of $U_{p_d,\alpha}$ is equal to $U_{p_d,\alpha}$, i.e.,*

$$U_{p_d,\alpha} = \overline{\operatorname{conv}\{u \in U_{p_d,\alpha} \mid u \text{ extremal }\}}^{weak-*}$$

Proof. Corollary B.2, the Banach-Alaoglu Theorem and the weak-∗ continuity of S imply that $U_{p_d,\alpha}$ is compact with respect to the weak-∗ topology. Then the assertion is a direct application of the theorem of Krein-Milman; see, e.g., [3, Theorem 2.19]. □

Furthermore, if S is a finite rank operator, the extremal points can be characterized as follows (cf., e.g., [3, Proposition 2.177]).

Theorem B.4. *Suppose that $\dim \operatorname{Ran} S = N_S < \infty$. The extremal points of $U_{p_d,\alpha}$ can be written as a linear combinations of no more than N_S Dirac delta functions:*

$$\{u \in U_{p_d,\alpha} \mid u \text{ extremal }\} \subset \left\{ \sum_{j=1}^{N_S} \boldsymbol{u}_j \delta_{x_j} \;\middle|\; \boldsymbol{u}_j \in H_1,\ x_j \in D \right\}.$$

Proof. Let $u \in U_{p_d,\alpha}$ be extremal. The proof will be done by contradiction. Assume, therefore, that $\operatorname{supp} u$ consists of more than N_S points. Then, there exists a disjoint partition $\{D_n\}_{n=1,\ldots,N_S+1}$ of the set D with the properties

$$|u|(D_n) > 0 \quad \text{for all } n = 1, \ldots, N_S + 1.$$

Define for $n = 1, \dots, N_S + 1$ the restrictions

$$u_n = u|_{D_n} \in \mathcal{M}(D, H_1).$$

It is clear that $\|u_n\|_{\mathcal{M}(D,H_1)} = |u|(D_n) > 0$. Now, we consider the renormalized measures and their image under S, i.e.

$$v_n = \frac{u_n}{\|u_n\|_{\mathcal{M}(D,H_1)}},$$
$$w_n = Sv_n \in \operatorname{Ran} S \subset H_2,$$

and look for a nontrivial solution $\lambda \in \mathbb{R}^{N_S+1} \setminus \{0\}$ of the system of linear equations

$$\sum_{n=1}^{N_S+1} \lambda_n S v_n = \sum_{n=1}^{N_S+1} \lambda_n w_n = 0 \in \operatorname{Ran} S.$$

Since the number of equations is one smaller than the number of variables, such a solution exists. Without restriction, we may assume $\sum_{n=1,\dots,N_S+1} \lambda_n \geq 0$ (otherwise, we take the negative of λ). We define

$$\tau = \max_{n=1,\dots,N_S+1} \frac{|\lambda_n|}{\|u_n\|_{\mathcal{M}(D,H_1)}}$$

and u_+ and u_- as

$$u_\pm = u \pm \frac{1}{\tau} \sum_{n=1}^{N_S+1} \lambda_n v_n = \sum_{n=1}^{N_S+1} \left(1 \pm \frac{\lambda_n}{\tau \|u_n\|_{\mathcal{M}(D,H_1)}}\right) u_n.$$

Clearly, $u_+ \neq u_- \neq u$. By construction and linearity of S we have $Su_\pm = Su = \widehat{p}$. Furthermore, we directly verify that

$$\begin{aligned}
\|u_\pm\|_{\mathcal{M}(D,H_1)} &= \int_D \mathrm{d}|u_\pm| = \sum_{n=1}^{N_S+1} \int_{D_n} \mathrm{d}|u_\pm| \\
&= \sum_{n=1}^{N_S+1} \left(\|u_n\|_{\mathcal{M}(D,H_1)} \pm \frac{\lambda_n}{\tau} \right) = \|u\|_{\mathcal{M}(D,H_1)} \pm \frac{1}{\tau} \sum_{n=1}^{N_S+1} \lambda_n
\end{aligned}$$

since $|\lambda_n|/\tau \leq \|u_n\|_{\mathcal{M}(D,H_1)}$. Since $\sum_{n=1,\dots,N_S+1} \lambda_n \geq 0$ we have $\|u_-\|_{\mathcal{M}(D,H_1)} \leq \|u\|_{\mathcal{M}(D,H_1)}$, and u_- is an optimal solution of (1.3), i.e., $u_- \in U_{\alpha,p_d}$ (Corollary B.2). Moreover, we see that it must hold

$$\sum_{n=1}^{N_S+1} \lambda_n = 0,$$

since the norm cannot be strictly smaller, since $u \in U_{\alpha,p_d}$. It follows that also u_+ is optimal. We conclude the proof with the observation that

$$u = \frac{1}{2} u_+ + \frac{1}{2} u_-,$$

which contradicts the assumption that u is extremal in U_{α,p_d}. □

The given proof can be modified into a constructive procedure to remove excess points from the support of an existing (suboptimal) solution of (1.3).

Proposition B.5. *Suppose that* $\dim \operatorname{Ran} S = N_S < \infty$. *Let* $u = \sum_{n=1,\dots,P} \boldsymbol{u}_n \delta_{x_n}$ *be a arbitrary with* $P \in \mathbb{N}$, $u_n \in H_1$, $x_n \in D$ *(pairwise distinct). Then, there exists a* $u^{new} = \sum_{n=1,\dots,P} \boldsymbol{u}_n^{new} \delta_{x_n}$ *such that*

$$\|u^{new}\|_{\mathcal{M}(D,H_1)} \le \|u\|_{\mathcal{M}(D,H_1)}, \quad Su^{new} = Su,$$

and all but N_S *of the coefficients* $\boldsymbol{u}_n^{new}$ *are equal to zero.*

Proof. The proof is done by induction on P. We only perform the step $N_S + 1$ to N_S. As in the previous proof, we define

$$u_n = u|_{\{x_n\}} = \boldsymbol{u}_n \delta_{x_n}, \text{ and } w_n = S(\boldsymbol{v}_n \delta_{x_n}), \text{ where } \boldsymbol{v}_n = \frac{\boldsymbol{u}_n}{\|\boldsymbol{u}_n\|_{H_1}}.$$

We find the nontrivial solution of $\sum_{n=1,\dots,N_S+1} \lambda_n w_n = 0$ with $\sum_{n=1,\dots,N_S+1} \lambda_n \ge 0$. Now, in contrast to the previous proof, we set

$$\tau = \max_{n=1,\dots,N_S+1} \frac{\lambda_n}{\|\boldsymbol{u}_n\|_{H_1}} \ge 0.$$

We set

$$u_{new} = u - \frac{1}{\tau} \sum_{n=1}^{N_S+1} \lambda_n \boldsymbol{v}_n \delta_{x_n} = \sum_{n=1}^{N_S+1} \left(1 - \frac{\lambda_n}{\tau \|\boldsymbol{u}_n\|_{H_1}}\right) \boldsymbol{u}_n \delta_{x_n}$$

Thus, the coefficients of u^{new} are given as $\boldsymbol{u}_n^{new} = [1 - \lambda_n/(\tau\|\boldsymbol{u}_n\|_{H_1})]\boldsymbol{u}_n$. It holds that $\|u^{new}\|_{\mathcal{M}(D,H_1)} = \|u\|_{\mathcal{M}(D,H_1)} - \sum_{n=1,\dots,N_S+1} \lambda_n/\tau \le \|u\|_{\mathcal{M}(D,H_1)}$ since $\lambda_n/\tau \le \|u_n\|_{H_1}$ and we finish the proof with the observation that

$$\boldsymbol{u}_{\widehat{n}}^{new} = 0 \quad \text{for } \widehat{n} \in \operatorname*{arg\,max}_{n=1,\dots,N_S+1} \frac{\lambda_n}{\|\boldsymbol{u}_n\|_{H_1}}. \qquad \square$$

Appendix C. Weak-∗ convergence of discrete measures

We prove the closedness of sets comprising vector measures supported on a uniformly bounded number of support points with respect to the weak-∗ topology on $\mathcal{M}(D, H_1)$.

Proposition C.1. *Let* D *be compact. For any* $N_d \in \mathbb{N}$ *the set*

$$P^{N_d} = \left\{ \sum_{j=1}^{N_d} \boldsymbol{u}_j \delta_{x_j} \;\middle|\; \boldsymbol{u}_j \in H_1, x_j \in D \right\}$$

is weak-∗ closed.

Proof. Let an arbitrary weak-∗ convergent sequence $\{u_k\}_{k\in\mathbb{N}} \subset P^{N_d}$ with limit $\widehat{u}$ be given. For each $k \in \mathbb{N}$ there exist $\boldsymbol{u}_j^k \in H_1$, $x_j^k \in D$, $j = 1, \dots, N_d$ with

$$u_k = \sum_{j=1}^{N_d} \boldsymbol{u}_j^k \delta_{x_j^k} \quad \text{and} \quad \|u_k\|_{\mathcal{M}(D,H_1)} = \sum_{j=1,\dots,N_d} \|\boldsymbol{u}_j^k\|_{H_1} \le C,$$

for some $C > 0$. Introducing the vectors $\boldsymbol{u}^k = (\boldsymbol{u}_1^k, \dots, \boldsymbol{u}_{N_d}^k)^T \in H_1^{N_d}$ and $x^k = (x_1^k, \dots, x_{N_d}^k)^T \in D^{N_d}$, there exist a subsequence of $(\boldsymbol{u}^k, x^k) \in H_1^{N_d} \times D^{N_d}$ denoted by the same symbol and $(\boldsymbol{u}, x) \in H_1^{N_d} \times D^{N_d}$ with $\boldsymbol{u}^k \rightharpoonup^* \boldsymbol{u}$ and $x^k \to x$ due to the compactness of D and the boundedness of $\boldsymbol{u}^k$. Defining

$$u = \sum_{j=1,\dots,N_d} \boldsymbol{u}_j \delta_{x_j},$$

we arrive at

$$\langle \varphi, u \rangle = \lim_{k\to\infty} \sum_{j=1,\dots,N_d} (\boldsymbol{u}_j^k, \varphi(x_j^k))_{H_1} = \lim_{k\to\infty} \langle \varphi, u_k \rangle = \langle \varphi, \widehat{u} \rangle$$

for all $\varphi \in \mathcal{C}_0(D, H_1)$ since $\boldsymbol{u}_j^k \rightharpoonup \boldsymbol{u}_j$ and $\|\varphi(x_j^k) - \varphi(x_j)\|_{H_1} \to 0$. Due to the uniqueness of the weak-$*$ limit we get $\widehat{u} = u \in P^{N_d}$ yielding the weak-* closedness of P^{N_d}. □

As a corollary each accumulation point of a sequence of measures with uniformly bounded support size is also finitely supported.

Corollary C.2. *Let D be compact. Consider a sequence $u_k \in \mathcal{M}(D, H_1)$ with $\#\operatorname{supp}|u_k| \le N_d$ for some $N_d \in \mathbb{N}$. Then every accumulation point $\widehat{u}$ of u_k fulfills $\#\operatorname{supp}|\widehat{u}| \le N_d$.*

Proof. Since every measure of support less that N_d can be written as a sum over N_d Dirac delta functions (by possibly adding additional Dirac delta functions with zero coefficient), applying Proposition C.1 yields the result. □

References

[1] J.-M. Azaïs, Y. de Castro, and F. Gamboa, *Spike detection from inaccurate samplings*, Appl. Comput. Harmon. Anal., 38 (2015), pp. 177–195.

[2] A. Bermúdez, P. Gamallo, and R. Rodríguez, *Finite element methods in local active control of sound*, SIAM J. Control Optim., 43 (2004), pp. 437–465.

[3] J. F. Bonnans and A. Shapiro, *Perturbation analysis of optimization problems*, Springer Series in Operations Research, Springer, New York, 2000.

[4] N. Boyd, G. Schiebinger, and B. Recht, *The alternating descent conditional gradient method for sparse inverse problems*, SIAM J. Optim., 27 (2017), pp. 616–639.

[5] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, New York, NY, USA, 2004.

[6] K. Bredies and H. K. Pikkarainen, *Inverse problems in spaces of measures*, ESAIM Control Optim. Calc. Var., 19 (2013), pp. 190–218.

[7] M. Burger and S. Osher, *Convergence rates of convex variational regularization*, Inverse Problems, 20 (2004), pp. 1411–1421.

[8] E. J. Candès and C. Fernandez-Granda, *Super-resolution from noisy data*, J. Fourier Anal. Appl., 19 (2013), pp. 1229–1254.

[9] E. J. Candès and C. Fernandez-Granda, *Towards a mathematical theory of super-resolution*, Communications on Pure and Applied Mathematics, 67 (2014), pp. 906–956.

[10] E. Casas, C. Clason, and K. Kunisch, *Approximation of elliptic control problems in measure spaces with sparse solutions*, SIAM J. Control Optim., 50 (2012), pp. 1735–1752.

[11] E. Casas, B. Vexler, and E. Zuazua, *Sparse initial data identification for parabolic PDE and its finite element approximations*, Math. Control Relat. Fields, 5 (2015), pp. 377–399.

[12] D. Colton and R. Kress, *Inverse acoustic and electromagnetic scattering theory*, vol. 93 of Applied Mathematical Sciences, Springer, New York, third ed., 2013.

[13] M. Dauge, *Elliptic boundary value problems on corner domains*, vol. 1341 of Lecture Notes in Mathematics, Springer-Verlag, Berlin, 1988. Smoothness and asymptotics of solutions.

[14] Y. de Castro and F. Gamboa, *Exact reconstruction using Beurling minimal extrapolation*, J. Math. Anal. Appl., 395 (2012), pp. 336–354.

[15] I. Dokmanić and M. Vetterli, *Room helps: Acoustic localization with finite elements*, in 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), March 2012, pp. 2617–2620.

[16] J. Droniou, *Solving convection-diffusion equations with mixed, Neumann and Fourier boundary conditions and measures as data, by a duality method*, Adv. Differential Equations, 5 (2000), pp. 1341–1396.

[17] V. Duval and G. Peyré, *Exact support recovery for sparse spikes deconvolution*, Found. Comput. Math., 15 (2015), pp. 1315–1355.

[18] I. Ekeland and R. Témam, *Convex analysis and variational problems*, vol. 28 of Classics in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, english ed., 1999. Translated from the French.

[19] B. Engquist and A. Majda, *Absorbing boundary conditions for the numerical simulation of waves*, Math. Comp., 31 (1977), pp. 629–651.

[20] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, Springer, 2015.

[21] M. Grant and S. Boyd, *CVX: Matlab software for disciplined convex programming, version 2.1*. http://cvxr.com/cvx, Mar. 2014.

[22] J. A. Griepentrog and L. Recke, *Linear elliptic boundary value problems with non-smooth data: Normal solvability on Sobolev-Campanato spaces*, Math. Nachr., 225 (2001), pp. 39–74.

[23] P. Grisvard, *Elliptic Problems in Nonsmooth Domains*, vol. 24 of Monographs and Studies in Mathematics, Pitman (Advanced Publishing Program), Boston, MA, 1985.

[24] R. Herzog, G. Stadler, and G. Wachsmuth, *Directional sparsity in optimal control of partial differential equations*, SIAM J. Control Optim., 50 (2012), pp. 943–963.

[25] B. Hofmann, B. Kaltenbacher, C. Pöschl, and O. Scherzer, *A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators*, Inverse Problems, 23 (2007), pp. 987–1010.

[26] F. Ihlenburg, *Finite element analysis of acoustic scattering*, vol. 132 of Applied Mathematical Sciences, Springer-Verlag, New York, 1998.

[27] K. Kunisch, K. Pieper, and B. Vexler, *Measure valued directional sparsity for parabolic optimal control problems*, SIAM J. Control Optim., 52 (2014), pp. 3078–3108.

[28] K. Kunisch, P. Trautmann, and B. Vexler, *Optimal control of the undamped linear wave equation with measure valued controls*, SIAM Journal on Control and Optimization, 54 (2016), pp. 1212–1244.

[29] J.-L. Lions and E. Magenes, *Non-homogeneous boundary value problems and applications. Vol. I*, Springer, 1972. Die Grundlehren der mathematischen Wissenschaften, Band 181.

[30] A. Milzarek and M. Ulbrich, *A semismooth newton method with multidimensional filter globalization for l_1-optimization*, SIAM Journal on Optimization, 24 (2014), pp. 298–333.

[31] P. Quittner and P. Souplet, *Superlinear parabolic problems: blow-up, global existence and steady states*, Springer Science & Business Media, 2007.

[32] J. L. Roux, P. T. Boufounos, K. Kang, and J. R. Hershey, *Source localization in reverberant environments using sparse optimization*, in 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, May 2013, pp. 4310–4314.

[33] W. Rudin, *Functional analysis*, International Series in Pure and Applied Mathematics, McGraw-Hill Inc., New York, second ed., 1991.

[34] G. Schiebinger, E. Robeva, and B. Recht, *Superresolution without separation*, Information and Inference: A Journal of the IMA, 7 (2018), pp. 1–30.

[35] G. Stampacchia, *Le problème de Dirichlet pour les équations elliptiques du second ordre à coefficients discontinus*, Ann. Inst. Fourier (Grenoble), 15 (1965), pp. 189–258.

[36] T. Suzuki, *L_1 generalized inverse beam-forming algorithm resolving coherent/incoherent, distributed and multipole sources*, Journal of Sound and Vibration, 330 (2011), pp. 5835–5851.

[37] G. M. Troianiello, *Elliptic differential equations and obstacle problems*, The University Series in Mathematics, Plenum Press, New York, 1987.

[38] J. A. Tropp, *Greed is good: algorithmic results for sparse approximation*, IEEE Transactions on Information Theory, 50 (2004), pp. 2231–2242.

[39] B. D. V. Veen and K. M. Buckley, *Beamforming: a versatile approach to spatial filtering*, IEEE ASSP Magazine, 5 (1988), pp. 4–24.

[40] A. Xenaki and P. Gerstoft, *Grid-free compressive beamforming*, The Journal of the Acoustical Society of America, 137 (2015), pp. 1923–1935.

[41] A. Xenaki, P. Gerstoft, and K. Mosegaard, *Compressive beamforming*, The Journal of the Acoustical Society of America, 136 (2014), pp. 260–271.

Dept. of Scientific Computing, Florida State University, 400 Dirac Science Library, Tallahassee, FL 32306, USA
E-mail address: kpieper@fsu.edu

Institute of Mathematics and Scientific Computing, University of Graz, Heinrichstrasse 36, 8010 Graz, Austria
E-mail address: quoc.tang@uni-graz.at, philip.trautmann@uni-graz.at

Center for Mathematical Sciences, Chair M17, Technische Universität München, Boltzmannstr. 3, 85748 Garching bei München, Germany
E-mail address: walter@ma.tum.de